ERIODICAL ROOM
ENERAL LIBRARY
UNIV. OF MICH.

MAY 20 1926

The Journal of EDUCATIONAL PSYCHOLOGY

Devoted Primarily to the Scientific Study of Problems of Learning and Teaching.

BOARD OF EDITORS

HAROLD RUGG, *Chairman.*

J. CARLETON BELL — ARTHUR I. GATES — RUDOLF PINTNER — L. M. TERMAN
FRANK N. FREEMAN — V. A. C. HENMON — BEARDSLEY RUML — E. L. THORNDIKE

VOL. XVII.
No. 5

MAY, 1926

$4.00 a year.
60c per copy.

CONTENTS

METHODS AND THEORIES OF LEARNING TO SPELL TESTED BY STUDIES OF DEAF CHILDREN. *Arthur I. Gates and Esther Hemke Chase* 289

THE INTERPRETATION OF BURT'S REGRESSION EQUATION. *Godfrey H. Thomson*. 301

AN EMPIRICAL STUDY OF THE SPEARMAN-BROWN FORMULA AS APPLIED TO EDUCATIONAL TEST MATERIAL. *Giles M. Ruch, Luton Ackerson and Jesse D. Jackson* 309

MENTAL RETARDATION IN SCHOOL CHILDREN INFESTED WITH HOOKWORMS. *W. G. Smillie and Cassie R. Spencer*. 314

AN INVESTIGATION OF THREE PLANS FOR SELECTING THE STUDENTS TO BE ADMITTED TO COLLEGE. *Esther Crane* 322

A METHOD FOR EVALUATING A PSYCHOGRAPH FOR VOCATIONAL GUIDANCE. *J. B. Miner* 331

A NEW INVESTIGATION OF JUDGMENTS ON THE SEX OF HANDWRITING. *J. S. Kinder*. 341

NOTES ON ARTICLES IN EDUCATIONAL PSYCHOLOGY IN CURRENT ISSUES OF OTHER MAGAZINES 345

NEW PUBLICATIONS IN EDUCATIONAL PSYCHOLOGY AND RELATED FIELDS OF EDUCATION 348

Published Monthly Except June to August by

Warwick & York, Inc.,

York, Pa. — Baltimore, Md.

Entered as Second Class matter November 15, 1921, at the Post Office at York, Pennsylvania under the Act of March 3, 1879.

The Journal of Educational Psychology

is published monthly, by Warwick & York, Inc., Baltimore, Md., and York, Pa. Title page and index are bound in December number.

Manuscripts for publication, books or other materials for review, and news items should be addressed to Harold Rugg, Chairman of the Editorial Board, 400 W. 118th Street, New York City.

The price of the Journal is $4.00 a year, payable in advance. Foreign postage is 30 cents extra. Single issues are 60 cents each, and less than a full year's subscription is charged at single issue rate for every month ordered.

Subscribers should notify the publishers of change in address at least four weeks before publication of the issue with which change is to take effect. No claim for nonreceipt of an issue will be entertained unless made within two weeks after receipt of the next succeeding issue.

Warwick & York, Inc.,
Baltimore, Md.

TEACHERS' HEADQUARTERS

European
Plan

Entirely
Fire-Proof

Centrally
Located

EDWARD DAVIS
Manager

HOTEL RENNERT
BALTIMORE, MARYLAND

HARVARD UNIVERSITY

SUMMER SCHOOL OF ARTS AND SCIENCES AND OF EDUCATION
July 6 to *August* 14, 1926

THE PSYCHOLOGY OF MENTALLY DEFICIENT, RETARDED, AND PSYCHOPATHIC CHILDREN

Course for School Psychologists and Special Class Teachers

LECTURES, clinics, demonstration classes for the study of the methods of training given by Professor Walter F. Dearborn, Professor Edwin A. Shaw, with the coöperation of Dr. Ransom A. Greene, Superintendent of the Walter E. Fernald State School, Waverley, Dr. Karl M. Bowman, Chief of Staff, Boston Psychopathic Hospital, and of other specialists in this field.

For information concerning this course and other courses in Education and related subjects, address the
Secretary of the Summer School

HARVARD UNIVERSITY
Cambridge, Massachusetts

"I hail Bagley's book as the most important offering of the year. But you who are afraid that I boil over too easily, you who want to follow a master reasoner careful of his route, step by step, read Bagley on Determinism even through its last appendix." Educational Review, March, 1926.

DETERMINISM IN EDUCATION
By William C. Bagley

IN THIS book, the author reports in full the critical and constructive work that he has done during the past four years in connection with the fundamental questions raised by the efforts to measure native intelligence. Among the papers making up the volume are: Democracy and the IQ; Further Fallacies of Determinism; Do Good Schools Pay?; Education and the Moral Hazard; The Post-War Period as a Test of Universal Education; The Army Intelligence Tests and the Pro-Nordic Propaganda; Education as a Creative and an Equalizing Agency. The data presented in the studies heretofore published by the author have been greatly amplified and brought up to date. The recent investigations which have so signally confirmed the position that Mr. Bagley took in 1922 are described and summarized. An appendix includes ratings of the several states on a number of measures—such as economic efficiency, the production of leaders, intelligence, morality and criminality—as well as revised ratings of State school systems.

$2.20 + 10¢ *postage*

WARWICK AND YORK, Inc.
PUBLISHERS BALTIMORE

THE JOURNAL OF
EDUCATIONAL PSYCHOLOGY

Volume XVII | May, 1926 | Number 5

METHODS AND THEORIES OF LEARNING TO SPELL TESTED BY STUDIES OF DEAF CHILDREN

ARTHUR I. GATES

Professor of Education, Teachers College

ASSISTED BY

ESTHER HEMKE CHASE

University of Minnesota

A series of studies of the acquisition of various linguistic abilities by children congenitally deaf were begun two and a half years ago for the purpose of discovering certain facts which might be applied to the improvement of methods of teaching and learning of normal children. In this paper will be described the results of one of the preliminary studies and some of the plans for further investigations now under way or contemplated.

The first study, begun in November, 1923, was designed to throw some light on the influence of phonic[1] and phonetic experience and training upon reading and spelling. For several decades various good and ill effects have been attributed to these types of experience and training of which there are nearly innumerable varieties of formal and informal procedures Among those most competent to appraise these practices in the light of established principles of learning agreement does not, by any means, prevail. Nothing short of extensive experimental work is likely to yield a valid estimate of the merits of phonic and phonetic training. A system of teaching so widespread in use and

[1] By *phonic* is here meant any experience or training either in hearing or pronouncing words or both. Training in pronunciation, of course, involves experience in hearing one's voice. By *phonetic* is here meant training in either translating visual (printed or written) words or word elements—letters, syllables, phonograms—into sounds as in reading; or conversely, sounds into letters and letter combinations, as in spelling.

yet so varied in character as phonetics in America merits most careful investigation.

While developing plans for a comprehensive experimental attack upon the problem, it occurred to the writer that one method of discovering what abilities or deficiencies result from phonic and phonetic experiences would be to study, analytically, the achievements and performances of children who had had no such training. Children born deaf fulfill perfectly this requirement; having never heard sounds, they could not, of course, be given phonetic work of any sort.

Through the courtesy of Dr. Harris Taylor, Principal, and Miss Edith Buell, of the Institution for the Improvement of the Instruction of Deaf-mutes, of New York, children suitable for investigation were found. The status of these children is indicated below. The "Institution Grade" indicates the grouping within the school; the "Approximate School Grade" suggests a roughly equivalent level in the ordinary public school:

Institution grade	Approximate school grade	Number of pupils	Extreme ages	Average age
VIIB	IIIB	12	12–17	13.6
VIIIB	IV	11	13–16	14.2
IXB	VB	9	15–19	16.4
XIB	VIII	13	16–20	18.1

Since it was our purpose to compare the achievements of these pupils with those of normal children as nearly as possible equivalent by nature except in auditory capacity, a measure of general intelligence was sought. Of several instruments tried, the Pintner Non-language Tests seemed most satisfactory. The results of the measurements need not be given in detail since they indicate that the pupils studied were but slightly lower in intelligence than normal individuals of corresponding age.

Reading Ability of the Deaf

Before giving the tests of spelling ability, the achievements of the pupils in reading were ascertained since there are good reasons for believing that absolute spelling ability is greatly influenced by the level of ability and amount of experience in reading. Once the spelling of a word is learned it is probably kept alive in no small measure by being perceived repeatedly in the process of reading as well as by

experience in writing.[1] Spelling ability should, therefore, be evaluated in connection with the evidence concerning reading ability and experience.

The Burgess Reading Test was given to all pupils. The paragraphs of which this test is composed are approximately equal in difficulty to the following:

"These ducks are out walking on a cold winter day. The black duck is ahead; but she is unhappy because the cold snow hurts her feet. Make one short line under her feet to represent a board for her to stand on. Then make one more short line for the white duck to stand on. Now both ducks will keep their feet warm."

This material is easy reading for normal Grade III children and not too difficult for many in Grade II. The scores of the several grades on the test—which is five minutes in length—were as follows:

Institution grade	Average chronological age	Per cent making 0 score	Per cent reading 1 or 2 paragraphs	Per cent reading 3 or 4 paragraphs	Per cent reading 5 or more paragraphs	Average number of paragraphs	Average reading, age	Average reading, grade
VIIB	13.6	50	50	0	0	0.8	?	?
VIIIB	14.2	18	18	45	18	3.0	8.2	2.5
IXB	16.4	0	55	11	33	3.6	8.5	2.8
XIB	18.1	0	25	17	68	5.0	10.0	3.5

The results revealed in this table were the most surprising outcome of the study. Half of those in Grade VIIB (whose ages varied from 12 to 20 years) failed completely to read a single paragraph of this easy material in five minutes and the other half read but one or two passages. In the higher grades, the reading was usually so slow and laborious that it could be neither very enjoyable or profitable. How significant this deficiency is and what the theoretical possibilities of remedying it are will be discussed in a later paper. At present, we shall utilize the results merely to illuminate the achievements in spelling.

Spelling Ability of the Deaf

The test of spelling ability was a list of 36 words from the Ayres-Buckingham Scale. The list comprised nine groups of 3 words each

[1] For a defense of this belief see A. I. Gates: The Psychology of Reading and Spelling. New York: Teachers College, Bureau of Publications, 1922.

of the same length beginning with 2 letter words and culminating with 11-letter words. The nine groups of words, when applied to normal children, form a series of approximately equal steps of difficulty. Since an age scale has been constructed for scores on this test, it is possible to appraise the achievements of the deaf by translating their raw scores into the chronological age of normal children who average the same score. The results are given in the accompanying table:

Grade	Average actual age of deaf	Average spelling age
VIIB	13.6	11.2
VIIIB	14.2	11.6
IXB	16.4	12.6
XIB	18.1	13.0

While the absolute achievements of the deaf are not as high as those of normal children of corresponding ages, they are distinctly excellent when certain relevant factors are taken into account. First, the meagerness of the reading and other linguistic experiences of the deaf constitutes a highly potent handicap. Compared to their ability in reading, the deaf pupils' spelling achievements are very high. Secondly, the examination as given proved to be not only a test of spelling but also of lip-reading ability. Despite all efforts to make the dictations intelligible, many quite obviously misinterpreted the pronunciations. For example, among the "misspellings" of *estimate* appear the following: *examine, depend, expect, Easter,* etc. and a number of blanks which meant inability to catch the word. Such "misspellings" and omissions ran through all of the records. Such words were called wrong, with a few exceptions: *build* substituted for *built* in over half of the cases; *difficulty* for *difficult; teacher* for *teach* and a few similar spellings were counted correct. These allowances, however, scarcely provided equal spelling opportunity for the deaf. Finally, some of the words appeared to be utterly unknown to the deaf. For example, *information* was attempted by only three pupils in the entire group and of these only one apparently had any idea what the word was. Unfamiliarity with abstract words of this type was very evident in the deaf children's reading and was doubtless in no small measure responsible for their difficulty in this subject.

Since it has been established that reading experience and as a consequence, reading ability exert a pronounced influence on spelling it will be illuminating and probably not unfair to compare the spelling of deaf children with that of normal children *of the same reading ability*. This may be done in terms of the age scale for normal children. The results are as follows:

Deaf group	Spelling age of deaf children	Spelling age of normal children of same reading ability	Superiority of deaf, in years
VIIB	11.2	6.9	4.3
VIIIB	11.6	8.1	3.5
IXB	12.6	8.4	4.2
XIB	13.0	9.4	3.6

The table shows that, with reading ability equal, the deaf greatly excel the normal children in spelling ability; the differences average but a trifle less than four years on a spelling age scale based on normal children. The spelling ability of the deaf, thus computed, is about 150 per cent that of the normal child.

To appraise in another way the spelling ability of the deaf, a further test was devised. This was termed a "spelling recognition" test. It consisted of 36 words ranging in length and difficulty, each presented with four misspellings, as in the following samples:

7. nime, nin, nyme, nine, nein

20. circular, sircular, circlear, cirkilar, sirculer.

36. consensions, kontientions, conscientious, konsienscious, constientious.

The task was to underline the correct word. Neither for deaf nor normal pupils were the words pronounced.

In this test the deaf children in the lower grades did relatively poorly—mainly I judge because they were quite unfamiliar with many of the words in speech, writing or print—but relatively well in the upper grades. In the highest group all scored 31 or more correct out of 36 and averaged 33.8, a figure which is about equal to the score of normal pupils in the seventh grade at midyear.

When the deaf children are compared in this test with normal pupils of equal reading ability, the results are as follows:

Deaf group	Deaf spelling recognition age	Normal subjects of same reading ability	Difference in favor of deaf in years
VIIB	9.4	6.9	2.5
VIIIB	10.8	8.1	2.7
IXB	12.7	8.4	4.3
XIB	14.6	9.4	5.2

The superiority of the deaf, when reading ability is rendered constant, is very marked and increases perceptibly with longer school experience. If these data are reliable and valid, the deaf child's ability to distinguish correct from incorrect spellings increases with extraordinary rapidity, in comparison with normal children, as their reading experience is extended.

In order to avoid the handicaps of the deaf due to unfamiliarity with abstract words and the earlier difficulty of conveying them the word to be spelled, a test of 16 names of concrete objects was presented by means of the lips in combination with a drawing to suggest the proper word. The words were: *hatchet, bicycle, soldier, umbrella, elephant, scissors, children, policeman, cigarette, telephone, butterfly, automobile, typewriter, battleship, locomotive, phonograph.* The number of correct spellings (out of a possible 16) on these tests was as follows:

Grade	Number Words Correct
VIIB	12.1
VIIIB	12.7
IXB	13.6
XIB	15.2

Since no scores for representative normal children on this list of words were available the test was given to the pupils in Grades IV, VI, and VIII in one New York City school and to Grades VA, VB and VIB in another. The scores were as follows:

Grade	School A	School B	Average A and B	Scores of deaf in corresponding grades
IV	7.8	....	7.8	12.7
VA	....	8.6	8.6	
VB	....	9.8	9.8	13.6
VI	11.0	10.6	10.8	
VIII	13.6	....	13.6	15.2

While the deaf children are older chronologically than the normal children in the corresponding grade, they have had much less reading and other language experience. The results shown in this table suggest a real superiority of the deaf.

On the whole, the evidence indicates that the deaf children possess, in comparison with their other liguistic abilities and in comparison with the achievements of normal children of similar reading experience, at least, superior spelling ability. If the sources of this superiority can be discovered, facts may be revealed which may make possible improvements in teaching spelling to normal children.

High spelling ability might be the result of extensive reading, extensive writing composition, extensive drill in spelling or of specially effective methods of learning and teaching in spelling. In this instance, it is certainly not due to extensive reading or writing (composition). It is quite surely not due to the amount of drill in spelling specifically, since except for occasional dictations of word lists, little time is devoted specially to spelling lessons. The means of teaching the pupils to learn to spell words, therefore, merit careful examination.

The spelling tests to which specific instruction in spelling is largely limited are of two forms: (1) mere requests to write words that occur to the pupils, (2) spelling words dictated by speech (lip-movements). In the first type of test no special teaching is used; in the second, a form of spelling technique is suggested. First the child tries to get the word by lip-reading and demonstrates his perception by pronunciation. This, in fact, is the way new words are introduced. The teacher speaks the word at the same time demonstrating its meaning by action or object. The pupils then try to pronounce the same word. If they cannot get it as a whole, they are assisted time after time in approximating the pronunciation of the word syllable by syllable, until they can reproduce the syllables in order perfectly or nearly so. Then the pupil writes the word as nearly as he can, usually syllable by syllable, on the board. When errors occur, the teacher gives the correct spelling by writing on the board the word with which the pupils compare their own written forms. It would seem that this method of developing the abilty to lip read, speak, spell and write all new words would tend to set up spelling habits not unlike those which prevail among normal children, namely, the device of relying primarily upon phonetic translation but closer examination may reveal significant differences. Normal children, enjoying phonic and phonetic experience, learn to depend primarily upon a phonetic rendering—translation

of the sounds into letters which represent them—when attempting to recall the spelling of words. In several analytic studies, published[1] and unpublished, of good and poor spellers this habit was found to be almost universal. That this method alone should function unsatisfactorily is due to the fact that English spelling is not systematically phonetic in the sense that a given sound is invariably represented by the same letters. *Claim* may be phonetically spelled as *clam, clame, klam, klame, claim, klaim* etc. The majority of spelling errors among normal persons are not due to erroneous phonetic rendering but to the inability to select the form which happens to be the traditional one from the many equally phonetic equivalents. In other words, by relying upon the device of simply translating sounds into a series of letters which does represent them, they are prone to error because so many possible translations are phonetically accurate.

The dictation lesson for the deaf encourages a habit of translating the lip-movement equivalents of syllables—which to the normal would be phonic units—into letters. The same difficulty prevails here as in phonetic translation, namely, the existence of many letter combinations which have been associated with the same lip-movements. But unlike the deaf, normal children have a life-time of phonic experience with language; long before they can read or write or spell they can recognize and articulate the word sounds. It is probably the greater facility in recalling the sounds and articulation of words that results in the common tendency to resort to utilizing the sound equivalents in spelling. The deaf, of course, are incapable of thinking first of the sounds and then recalling a combination of letters which represent them and despite the system of encouraging them to spell by recalling combinations of letters associated with the lip-movements of others or their own acts of pronunciation, it seems probable—both from observing the deaf during spelling and learning tests and from study of their misspellings—that they depend mainly on another learning device. This device, I believe, consists in a more careful visual study of the word forms in the final stages of learning when their written word is compared with the correct form placed on the board and in the attempt to recall not the lip-movement-letter combination associations but the visual appearance of the word during attempts to spell words not yet firmly habituated as writing habits. This hypothesis we attempted to test by further examinations.

[1] Gates: *Op. cit.*

The first procedure consisted in comparing the abilities of deaf and normal children to learn to spell words misspelled previously in a test. Immediately after completing the test on the 16 words—*hatchet, bicycle*, etc.—these words in mimeographed form were distributed to the pupils who were instructed to study them for another test to be given a minute later. They were required to study by observing the printed words; but they were not permitted to spell the words aloud or to write them. After a minute of study, the lists were collected and the words again dictated. In the accompanying table are given the results for 23 deaf pupils each paired with a normal child of identical score on the first test.

Score for Normal Pupils

Average initial score	Average final score	Average gain	PE gain	PE difference
11.5	13.9	2.4	0.18	

Score for Deaf Pupils

11.5	15.5	4.0	0.15	0.23

That the deaf learned more during the one minute period of study is apparent in the table. They acquired the spelling of 4 words while the normal learned 2.4. The difference between the gains (1.6 words) is approximately seven times the probable error of the difference. This superiority, under the conditions of the test, is in my judgment evidence that the deaf have a sharp perception for word-forms which is utilized in learning to spell. Another test of this hypothesis was also applied.

The "Word Perception" tests used considerably heretofore in the diagnosis of difficulties in reading and spelling among normal children were applied to the deaf. Both tests were designed to measure the speed and accuracy of perceiving words. The first, known as the Word Selection Test, measured the speed and accuracy with which a person can perceive an isolated word and immediately perceive it again in the midst of several other very similar (hence confusing) words. Two sample exercises from the test follow:

hat	sat	pot	rat	bat	hat
magnificent	mountainous distinction magnifficant	everlasting magnificent illustrious	magnifersant merchandise measurement	appointment maggnifsunt composition	

The second test, known as the Word Perception, Same-different Test, requires the subject to underline the pairs of words which are unlike leaving unmarked those identical, as in the following samples:

it	it	ground	grourd	children	chuldren
as	ar	behind	bekind	mountain	mountain
am	am	butter	butter	remember	remember
he	ha	rushed	rushed	thousand	thosand
me	me	marble	mardle	question	quastion

The tests obviously do not measure word-perception in exactly the same way but according to available evidence they are both symptomatic of the kind of perceptive skill which we desire here to measure. Since the tests correlated highly with familiarity with the words—that is, with reading experience and skill in which the normal and deaf differ widely—the two groups should be equated for reading ability before their achievements in this task are compared. Instead of presenting the data in this form, however, we shall show the scores in reading, spelling and word-perception since from these figures the significant facts may be estimated together with certain others. The scores below show the age of average normal children which the deaf equal in the four functions.

Deaf group	Actual age	Reading age	Spelling age	Word selection age	Word perception same-different
1	13.6	6.9[1]	11.2	12.4	16.5
2	14.2	8.1	11.6	12.4	16.0
3	16.4	8.4	12.6	15.5	17.7

[1] Approximately.

The results are as striking as they are unequivocal. The deaf Group 1, for example, which could scarcely read at all and which, consequently, could have enjoyed but little reading experience shows a word perception ability in one tests as great as that of average 16.5

year normal children, that is, normal children approximately three years above their own average chronological age. The results for the other groups were similar; Group 4 was omitted because their word perception ability exceeded all available norms for hearing children. There can be no doubt about the extraordinary word-perception ability of the deaf pupils.

Spelling ability, it may be observed, falls about midway between reading and word perception. Lack of reading experience tends to pull spelling ability down whereas keenness of word-perception tends to lift it up. This, at least, is our interpretation. The deaf owe their remarkable spelling ability primarily to a peculiarly effective type of perceiving, of reacting visually to words. Normal children, despite wider opportunities in reading and writing, do not seem to acquire as effective a type of word-perception. The hypothesis here offered is that normal children fail to develop this precise, accurate and—as far as the effects on spelling are concerned—effective form of word observation because they rely mainly on the easier, perhaps more natural—yet for spelling less productive—device of phonetic translation.

The results of the study of spelling by the deaf contribute certain suggestions for improving spelling ability among normal children. Practice in spelling may be directed to give primary emphasis to:

1. The motor activities of articulating the constituent letters (or letter-units such as syllables) silently or aloud.
2. The motor activities of writing the letters.
3. The mental activities of translating recalled word-sounds into letters which may be imaged, articulated or written.
4. The visual-perceptive activities of "seeing" the visual word as a whole of clearly distinctive parts.
5. The recall and application of spelling or phonetic rules which, in combination with (3) assist the learner to select from the many possible the one correct phonetic rendering.

Studies of spelling already presented and others to be published soon indicate that normal pupils rely greatly upon (1) and (3). The success of the deaf indicates that much may be gained by establishing better habits of visual study of words in learning to spell and then of exercising until firmly mastered the habit of writing the word. Studies to be reported indicate that elementary school pupils write or articulate the spelling only after recalling the letters by translation of sounds to letters or by visualization of letters or by both combined. Since English is so unphonetic, the most reliable device for deter-

mining which of the many genuine sound-to-letters translation is the correct one is to be able to recall some impression of the visual appearance of the word. This, I believe, is the method adopted by the deaf—with excellent results as we have seen. While the use of spelling rules has largely fallen into disrepute there are theoretical reasons and some experimental evidence[1] tending to show that when used as one of the means of emphasizing the visible similarities and differences among groups of like and unlike word forms they may be fruitfully employed to improve word-perceptive habits and, consequently, spelling ability.

[1] Miss Alice Watson: Experimental Studies in the Psychology and Pedagogy of Spelling. Teachers College, Contribution to Education, forthcoming number.

THE INTERPRETATION OF BURT'S REGRESSION EQUATION

GODFREY H. THOMSON

Moray House, University of Edinburgh

In this *Journal* for December, 1925, Holzinger and Freeman question the accuracy of the interpretation which Burt has placed upon a regression equation worked out by him in *Mental and Scholastic Tests, viz.:*

$$\text{Binet} = .54\ \text{School Work} + .33\ \text{Intelligence} + .11\ \text{Age}.$$

They further discount very largely, if they do not deny, an argument which I have sketched in a recent book[1] to the effect that Burt's Reasoning Test is a measure of native intelligence entirely independent of schooling.

The object of the present note is to make my own position quite clear in the matter, for while agreeing with a proper expression of what Burt deduces from the above-mentioned experiment, I am also very fully aware of the dangers of regarding correlations as proofs of structural relationships. In another connection, also involving correlational relationships, I have indeed rather warmly championed the very point which Holzinger and Freeman here emphasize.

Part of the trouble is due to the loose use of terms, to which one is continually tempted because pedantic exactitude is repellent to any audience or to any reader. Nevertheless, it was unpardonable of me to use "schooling" in the two senses indicated by Holzinger and Freeman, in successive paragraphs of my book. The one sense is that of length of schooling: the other is that of school achievement in which, of course, two boys may differ even if they have the same schooling in the former sense. The correction of this ambiguity on pages 217 and 218 of my book would modify the expression of the conclusions, but would, I think, not destroy them.

This ambiguity could have been avoided, and I regret it. Some of the ambiguities in Burt's words are however very difficult to avoid. The use of such terms as "attributable to," "is caused by," etc. is almost inevitable except among those technically proficient in the notation and symbolism of correlation. As Holzinger and Freeman rightly and forcibly point out, this usage would lead, if persisted in, to such absurdities as saying that over half of a child's age is "attributable

[1] *Instinct, Intelligence and Character.* Longmans, Green and Co., 1925, pp. 217.

to" school achievement. Holzinger and Freeman themselves, in their last paragraph, give an accurate expression of the meaning of regression coefficients. "The regression coefficients," they say, "show merely the average change in the dependent variable for a unit change in the independent variable to which they are attached, the remaining variables being held constant." An attempt to expound in greater detail the meaning of such coefficients will be made below.

"To interpret them as representing the parts or portions of the independent variables which go to make up the predicted variable is," they go on to say, "unwarranted." And they are strictly correct. As they further point out, it is only by a coincidence that the three coefficients in Burt's regression equation happen to add up to 0.98, thus adding to the plausibility of the naive explanation. This I have always pointed out when lecturing on this subject, and have sometimes been on the verge of saying it in print, but have always reflected that it was difficult to do so without appearing to disagree entirely with Burt's conclusion.

A general audience is always very puzzled by regression equations, and receives with a disconcerted air the apparently paradoxical fact that, for instance,

$$\text{Binet} = .54\ \text{School Work} + .33\ \text{Burt} + .11\ \text{Age}$$

and *simultaneously*[1]

$$\text{School Work} = .69\ \text{Binet} - .05\ \text{Burt} + .47\ \text{Age}$$
$$\text{Burt} = .95\ \text{Binet} - .10\ \text{School Work} + .07\ \text{Age}$$
$$\text{Age} = .15\ \text{Binet} + .51\ \text{School Work} + .03\ \text{Burt}$$

The idea of estimating the most probable value of one variable when the values of certain others are known is that which can most readily be employed without error. The first of these equations means that, in the absence of a measured Binet Age for a given subject, the most probable value is that obtained by taking .54 School Achievement Age + .33 Burt's Reasoning + .11 Age, all quantities being measured from their average value in the population in question, and in *sigma* units. Individual estimates of Binet Age made by this means would not always be true, but such individual estimates would as often overshoot

[1] These equations I quote from Holzinger and Freeman's calculation. As they surmise, I had made similar calculations while making the reflections which led me to mention Burt's equation in my book, and except for one second place discrepancy my figures agree with theirs.

the mark as fall short of it. The "Binet" on the left hand side of that first equation is not an actual but a probable value: or if a number of cases happened to coincide in the other three quantities, that "Binet" would be the average of their Binet Ages. In the other equations, where Binet occurs on the right hand side, it means an actual measured Binet Age.

Burt's Regression Equation, as the first of these four equations has come to be called, says that in estimating a child's Binet Age without measuring it, School Work is the most important factor. The Binet Age fluctuates more markedly with School Work, other things being equal, than it does with Burt's Reasoning Age (Intelligence, as he calls it) other things again being equal.

The equation which Holzinger and Freeman have very cleverly chosen to enable them to poke fun at Burt's rhetoric is the fourth. It only says, however, that age, if unknown to us, can be estimated from Binet, School Work, and Burt, and that School Work is much the most useful of these in enabling us to form a true estimate of the unknown age. Other things being equal, Age fluctuates markedly with School Work. We do not of course say that Age is "attributable to" School Work. We do not put the cart before the horse. But the cart and the horse are causally connected, and if my cart and horse are missing, knowledge of the whereabouts of the cart will assist me in finding the horse. If we notice a coincidence in the times of full moon and of high tide, we are not justified in saying that the high tide causes a full moon. Yet we do say, *from the same observation*, that the full moon causes a high tide. The fact that it is absurd, from certain data, to draw the conclusion that Age is "attributable to" School Work does not carry with it a necessary condemnation of him who from the same data concludes that Binet Age is largely "attributable to" School Work. The one conclusion states a conceivable connection, the other an obvious absurdity. But this absurdity does not overturn that possible connection. It is unwise to use the words "attributable to," no doubt, since the exact direction and composition of the underlying complex of causes is unknown. But it is quite impossible to avoid the inner conclusion that the most probable hypothesis here is simply that School Work assists the child to answer Binet questions, and a very heroic pedantry is required by anyone who would refrain entirely from saying so.

If we turn to the third equation, we see that School Work is of hardly any assistance in estimating how well a child will do in Burt's Reasoning Test. No one is more clearly aware than I am that there

are many underlying structures of conceivable causal connections which will explain the present data. But an obvious working hypothesis ready to hand is that in Burt's Reasoning Tests native intelligence alone is sufficient, and that when this is equal between two children of the same age, the degree of attainments in school does not appreciably influence the Burt score.

It will I hope be clear that I agree with Holzinger and Freeman in their contention that it is in all strictness wrong to draw conclusions from correlation coefficients as to underlying causes. But unless we are to give up all guidance in this work, we must take some causal hypothesis which is in accord with the correlational data and also with other general facts, and use it as an aid to thought, and as a means of communication with laymen.

Let us turn from psychological experiment to data where we actually know what is occurring, and see what a set of regression equations numerically identical with Burt's would there involve. Fig. 1 is a diagram of an arrangement of coin throws which gives Burt's correlations exactly.

Two hundred seventy coins are thrown. The artificial Binet score is to consist of the number of heads in 178 of these; the artificial Burt score of the number of heads in 161 of them; the artificial school score of the heads in 197; the artificial age of the heads in 235 of them. By the use of suitable units and constants, the quantities might even be made to ape actual ages. One hundred thirty-four of the coins, suitably marked, are to be counted in to each of the four variables. Others are specific to one variable, or are common to two or to three. For example 19 are specific to Burt, 34 are common to Age, School, and Binet, and so on. The intercorrelations which would be produced by a large number of throws, of 270 coins each time, with this structural arrangement would be:

Correlation of Binet with Burt,

$$= \frac{6 + 2 + 134}{\sqrt{178 \times 161}} = 0.84;$$

Burt with Age,

$$= \frac{2 + 134}{\sqrt{161 \times 235}} = 0.70;$$

Age with School,

$$= \frac{19 + 34 + 134}{\sqrt{235 \times 197}} = 0.87;$$

School with Binet,

$$= \frac{2 + 34 + 134}{\sqrt{197 \times 178}} = 0.91;$$

Binet with Age,

$$= \frac{2 + 34 + 134}{\sqrt{178 \times 235}} = 0.83;$$

School with Burt,

$$= \frac{134}{\sqrt{197 \times 161}} = 0.75.$$

The correlations are the same as in Burt's experiment, and the partial correlations of course follow. We might proceed to draw conclusions as to the underlying causal relationships of the coin throws here called Binet Tests, School Work, etc.: but if we did so we would have to be very cautious. For Fig. 2 is another arrangement of 270 coins, *which gives exactly the same correlations*, though the causes are different.

The correlations can be calculated as before, by the formula:—

$$r = \frac{\text{Number of common elements}}{\text{Geometrical mean of the two totals}}$$

They are again the same as Burt's. Figure 3 is a third such arrangement, this time with exactly 100 coins in each variable.

Again it gives Burt's correlations. And one could make an infinite number.[1] Nevertheless, in making the attempt, it becomes borne in upon one that certain structures are impossible with this set of correlations. Try to obliterate entirely the general factor, and this, that or the other group factor, and you soon learn that some are essential, within variable limits.

Although innumerable structures, then, can explain a given set of correlation coefficients, yet these structures vary only between limits, and of the limits, were our mathematics competent, something could be said. In any case, to return to our beginning, we can state, independently of structure, which variables are most potent in enabling us to predict others. When we are given three of the four quantities here under discussion, *School Work is more potent than any other of these factors in enabling us to estimate Binet Age: It is of practically no use in estimating Burt Age under such circumstances, if these data are confirmed.*

[1] These structural figures are formed with equivalent, all-or-none, atoms. Abandon this, and other structures again can be formed.

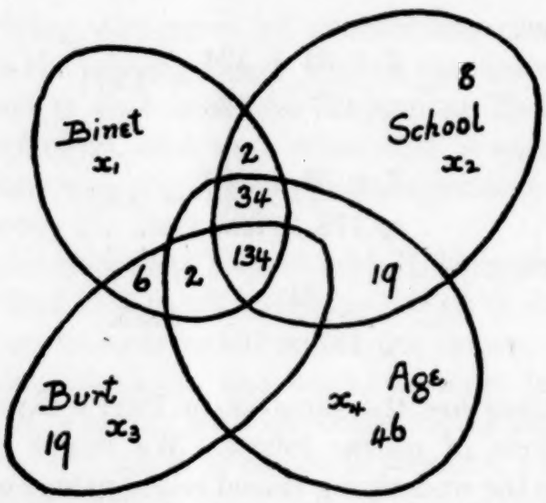


FIG. 1.—Apart from "general intelligence" (the 134 in the middle) *Burt* has here nothing in common with *School*, whereas *Binet* has much in common.

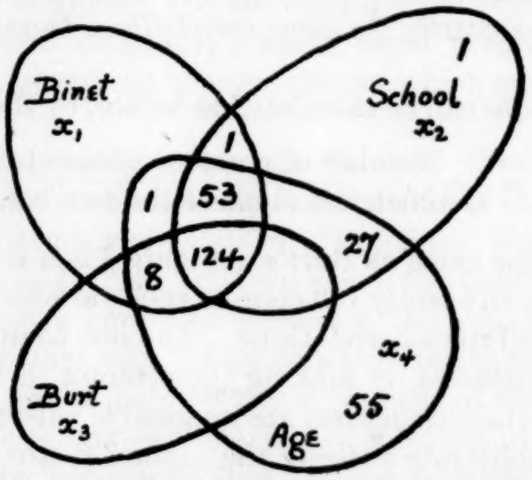


FIG. 2.—Here again *Binet* is, while *Burt* is not, linked to *School*, in addition to the "general intelligence" factor (124) in the center.

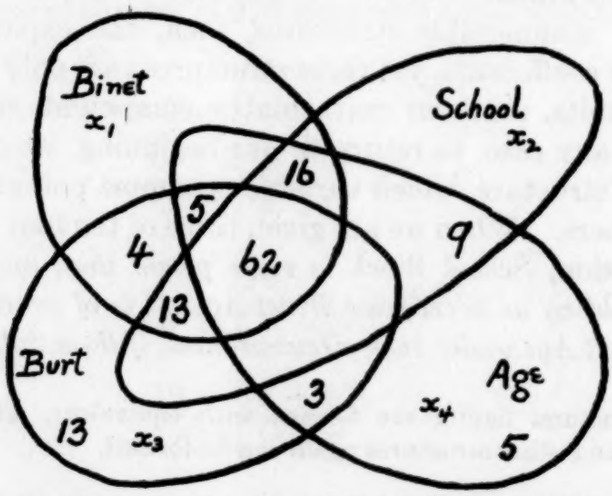


FIG. 3.—Here however, *Burt* as well as *Binet* is linked specially to *School*, independently of the central factor: but *Binet* has 16 + 13 such links, Burt only 13.

Appendix

It may be of some interest to note how, from these structure diagrams, the correlation of any given weighted team of three of the variables with the remaining variable can be found. For example we might find the correlation, with Binet, of

$$54 \text{ School} + 33 \text{ Burt} + 11 \text{ Age},$$

a correlation which, as the regression equation informs us, will be higher than that of any other combination of these variables with Binet. The ordinary formula for multiple correlation tells us that this correlation R is given by

$$(1 - R^2) = (1 - r^2_{12})(1 - r^2_{13.2})(1 - r^2_{14.23})$$

where x_1 = Binet, x_2 = School, x_3 = Burt, and x_4 = Age, and the correlations are those given by Burt, or in the present paper. There results

$$R = 0.94.$$

To show how this also can be deduced from the structure diagrams, certain additions must be made to the basic formula

$$r = \frac{\text{Number of common elements}}{\text{Geometrical mean of the two totals}}$$

For in a team, certain of the elements are repeated, since they occur in each of the variables, and also because the weights are equivalent to repetitions. The general principle is, that if a number of elements is repeated k times, that number is multiplied by k^2 in the denominator of the above fraction, but only by k in the numerator. To simplify the arithmetic, let us replace Burt's equation by

$$\begin{array}{c} 5 \text{ School} + 3 \text{ Burt} + \text{Age, } i.e., \\ 5x_2 \quad + \quad 3x_3 \quad + \quad x_4. \end{array}$$

Then using the structure shown in Fig. 1 the correlation between this team and x_1 will be

$$\frac{5 \times 2 + 3 \times 6 + 6 \times 34 + 4 \times 2 + 9 \times 134}{\sqrt{178}\,\sqrt{(25 \times 8 + 9 \times 19 + 46 + 36 \times 19 + 25 \times 2 + 9 \times 6 + 36 \times 34 + 16 \times 2 + 81 \times 134)}}$$

$= 0.94$ just as by multiple correlation. In this complicated looking but really very simple formula the 5×2 at the beginning of the numerator, for example, represents those two elements just above the

number 34 in the diagram. They are common to x_1, and to the team $5x_2 + 3x_3 + x_4$. In the team they are part of x_2 only, and their influence in the numerator is fivefold because x_2 is used fivefold, *i.e.*, has the coefficient 5 in the regression equation. Later in the numerator come another two elements, in the term 4×2. These two elements are those in the diagram between the numbers 6 and 134. They are part of x_1 (on the one hand) and of the team (on the other hand). In the team they occur both in x_3 and in x_4, threefold in x_3 and singly in x_4, and their coefficient in the numerator is therefore $3 + 1 = 4$. Similarly the reader can follow the treatment of other elements in the numerator.

In the denominator the number under the first square root is the total number of elements in x_1. The number under the second square root is composed of the total number of elements in the team, but any part is multiplied by the square of its frequency. For example, the first term is 25×8. The 8 stands for the 8 specific elements in x_2 (School) in the diagram, and it is multipled by 5^2 because 5 is the weight of x_2 in the regression equation. The last term is 81×134. It represents those 134 elements in the middle of the diagram which are common to all the variables, and the 81 is the square of $5 + 3 + 1$, the sum of the coefficients of the equation. It is only arithmetical clumsiness which blurs what is really a very simple principle.

Should the reader care, he may in similar fashion investigate the other structure diagrams, with a similar result.

AN EMPIRICAL STUDY OF THE SPEARMAN-BROWN FORMULA AS APPLIED TO EDUCATIONAL TEST MATERIAL[1]

GILES M. RUCH, LUTON ACKERSON AND JESSE D. JACKSON
University of Iowa

The theoretical formula,

$$r_{nn} = \frac{nr_{11}}{1 + (n-1)r_{11}},$$

whereby one can estimate the "reliability correlation" (r_{nn}) of n against n applications of the same or equivalent unitary tests from the reliability correlation (r_{11}) obtained from the application of a single pair of unit tests was derived independently by C. Spearman and W. Brown, and was published simultaneously by them in the same number of the *British Journal of Psychology*, Vol. III, October, 1910, p. 281 and p. 299. In its simple derivation by Brown the only assumptions required are that the standard deviations of the unitary tests must be equal and that the intercorrelations of all pairs of unitary tests must be equal. If these conditions were completely fulfilled in psychological or educational test data, the theoretical results obtained from the Spearman-Brown formula would by necessity coincide exactly with those obtained from amalgamations of repeated givings of the same or equivalent unitary tests. Since these conditions are perhaps never ideally fulfilled in practice, our problem may be stated: "How closely do the theoretical results prophesied by the formula accord with actually obtained results under the given conditions of inequality of unitary standard deviations and of intercorrelations?"

The test material of this experiment was carefully constructed as follows: A series of 500 words was taken from the lists of words correctly spelled by 97 per cent, 96 per cent, 95 per cent, or 94 per cent of unselected Grade VIII pupils as reported in the Iowa Spelling Scales (E. J. Ashbaugh, *Journal of Educational Research Monographs*, No. 3, 1922, pp. 131–134); all homonymous and cognate words being excluded. From these, 20 lists of 25 words each were drawn up so that each list contained the same number of words of equal difficulties. These lists were given in fixed order as "spelling exercises" to 500 pupils in Grades IVA, VB, VA, and VIB who had not formally been taught these lists.

[1] The writers wish to express their indebtedness to Dr. Paul McKee of the Hibbing, Minnesota, schools for the opportunity of this investigation.

The means and standard deviations of the 20 unitary tests were as follows:

Test	Mean	SD	Test	Mean	SD
1	16.958	5.116	11	15.158	5.450
2	19.056	4.530	12	16.998	5.402
3	16.656	5.291	13	16.336	5.750
4	16.672	4.947	14	17.580	4.928
5	16.618	5.325	15	16.346	5.787
6	17.666	5.258	16	16.962	4.709
7	16.650	5.520	17	17.166	5.233
8	17.188	5.219	18	16.920	4.893
9	16.776	4.672	19	17.470	5.130
10	16.222	5.200	20	16.002	5.388

Out of the possible 190 intercorrelations of the 20 unitary tests the following representative 20 intercorrelations (product moment) were calculated:

Tests correlated	*r*	Tests correlated	*r*
1 *vs.* 2	.850 ± .008	1 *vs.* 11	.831 ± .009
3 *vs.* 4	.828 ± .009	2 *vs.* 12	.828 ± .009
5 *vs.* 6	.870 ± .007	3 *vs.* 13	.860 ± .008
7 *vs.* 8	.911 ± .005	4 *vs.* 14	.789 ± .011
9 *vs.* 10	.848 ± .008	5 *vs.* 15	.888 ± .006
11 *vs.* 12	.823 ± .010	6 *vs.* 16	.832 ± .009
13 *vs.* 14	.829 ± .009	7 *vs.* 17	.861 ± .008
15 *vs.* 16	.865 ± .008	8 *vs.* 18	.850 ± .008
17 *vs.* 18	.871 ± .007	9 *vs.* 19	.808 ± .010
19 *vs.* 20	.805 ± .011	10 *vs.* 20	.889 ± .006

One notes that in spite of the care in constructing the tests the SD's and the *r*'s are not equal. There is no evidence, however, of any significant uniform "secular" increase or decrease in the sizes of these constants with successive givings of the unitary tests.

The mean value of these 20 unitary intercorrelations was .847, and this value was employed as the single unitary reliability *r* upon which the formulaic predictions were based. The empirical correlations were

obtained by pooling successively the odds against the evens among the 20 tests, *viz.*, 1 *vs.* 2, 1 + 3 *vs.* 2 + 4, 1 + 3 + 5 *vs.* 2 + 4 + 6, and so on. In the following table, n represents the number of paired givings; *e.g.*, when $n = 3$, the r's are those of the pool of three unitary tests against three other unitary tests. The last column represents for each calculation the divergence of the predicted r from the obtained r in terms of the PE's of the latter.

n	Predicted correlation	Obtained correlation	Divergence in terms of PE's of obtained r's
1	(.8470)[1]	.850 ± .0084	(−0.36)
2	.9171	.9025 ± .0056	2.61
3	.9439	.9267 ± .0043	4.04
4	.9577	.9464 ± .0031	3.59
5	.9651	.9596 ± .0025	2.16
6	.9707	.9695 ± .0018	0.66
7	.9748	.9737 ± .0016	1.72
8	.9779	.9760 ± .0014	1.33
9	.9803	.9799 ± .0012	0.33
10	.9811	.9812 ± .0011	−0.09

[1] This value is the mean of 20 representative intercorrelations.

The results are portrayed graphically in the diagram on the followin page.

One notes that in only two or three of the nine predictions are the divergences "statistically significant." The amount of agreement between theoretical and obtained results depends entirely upon the degree of "equivalence" of the unitary SD's and r's (and thus indirectly upon the "homogeneity" of the test material), and the form of the curve for the obtained r's depends upon the chance fluctuations in these unitary constants with successive givings. Since the Spearman-Brown formula is but a special case of the later Spearman[1] sums-and-differences formula under the assumptions of equality noted above, it would be possible to reconstruct the whole story in detail by writing out the appropriate expansions of the sums-and-differences formula for successive cumulations. Because of the arithmetical labor involved, however, it is unlikely that anyone will attempt such an analysis.

[1] Correlations of Sums and Differences. *British Journal of Psychology*, Vol. V, 1913, p. 419.

Four similar experiments have been reported. In the earliest, Holzinger[1] employed Forms A and B of the Terman Group Test upon 135 pupils, pooling tests A1 *vs.* B1, A1 + A2 *vs.* B1 + B2, etc. The marked "lag" of the empirical values which he found during the latter half of the amalgamations could be explained by the fact that the unitary *r*'s for the second half of the test-elements were much lower

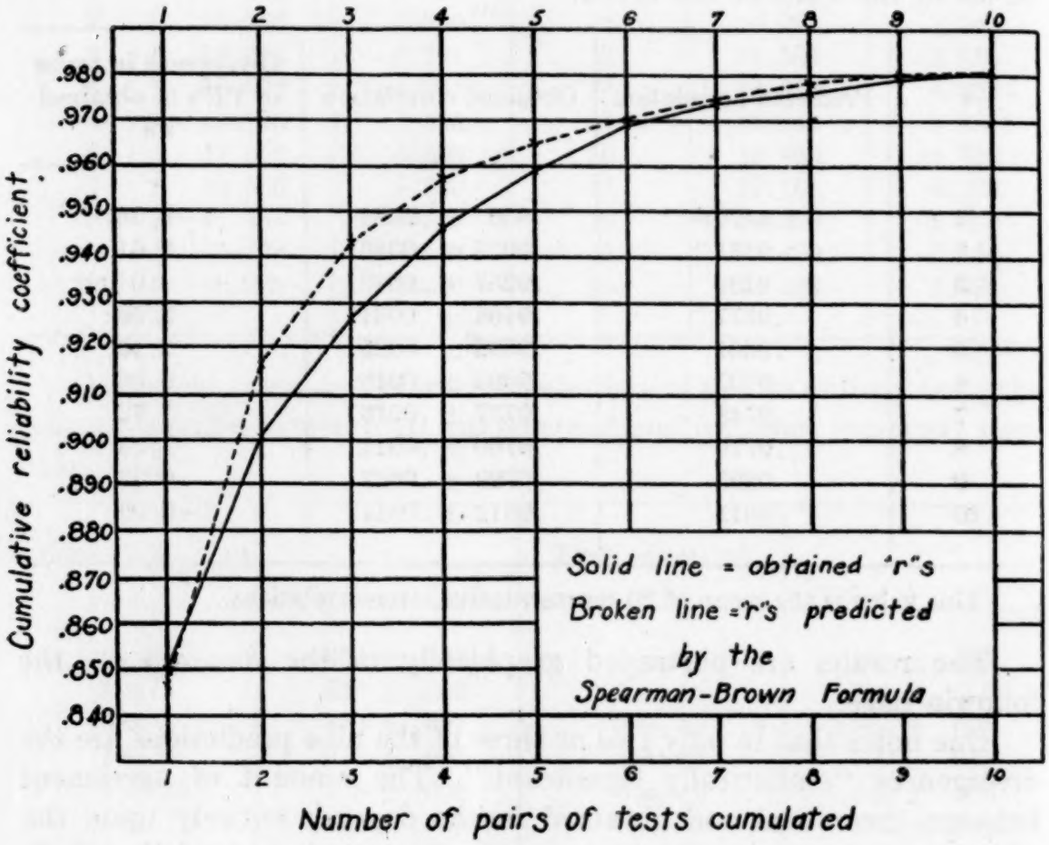


FIG. 1.

than those for the first half, unless by chance the SD's, which were not reported, should be of such magnitudes as fortuitously to compensate for this lowering tendency. Upon these sorts of data the Spearman-Brown formula could not be expected to apply. In two later experiments,[2] Holzinger and Clayton employed the scores made within 10 consecutive one-and-a-half-minute periods upon the two forms of the

[1] Note on the Use of Spearman's Prophecy Formula for Reliability. *Journal of Educational Psychology*, Vol. XIV, 1923, pp. 302–305.

[2] Further Experiments in the Application of Spearman's Prophecy Formula. *Journal of Educational Psychology*, Vol. XVI, 1925, pp. 289–299.

Otis Self-administering tests by about 100 graduate students. Here too the conditions of equality underlying the formula were not satisfied. However, by employing a "best" value of the unitary *r*'s ascertained by the method of least squares they found a close correspondence between theoretical and actual results. Upon other data obtained from homogeneous spelling lists given to 125 high school pupils they found that the theoretical coefficients based upon the *initial r* accorded closely with actual values. Such a close agreement, they note, is "probably accidental" since the initial *r* fell close to the "best" value. Kelley[1] upon the data of Gordon concerning the reliability (or perhaps more strictly the "objectivity") of lifted weights ascertained theoretically the average unitary intercorrelation of one judge's rankings against another's, and found that the predicted coefficients upon this value corresponded closely with obtained results.

In our present data the theoretical and actual coefficients tended to converge during the latter half of the cumulations, and ultimately become almost indistinguishable. This terminal convergence is probably due to the fact that the mean value of the 20 representative unitary *r*'s approached closely to a "best" value.

Conclusions

The conclusions seem to be (1) that in test material sufficiently homogeneous to induce fairly equivalent unitary SD's and *r*'s, or in case a "best" value of the unitary *r* is obtained, the Spearman-Brown formula gives a meaningful prediction, and (2) that out of five reported experiments significant divergences occurred in only two cases, and in these cases the tendency was for the formula to over-predict.

[1] The Applicability of the Spearman-Brown Formula for the Measurement of Reliability. *Journal of Educational Psychology*, Vol. XVI, 1925, pp. 300–303.

MENTAL RETARDATION IN SCHOOL CHILDREN INFESTED WITH HOOKWORMS[1]

W. G. SMILLIE, M.D.

AND

CASSIE R. SPENCER, M.A.

One of the most striking clinical symptoms of a severe case of hookworm disease is mental sluggishness. In a case of heavy hookworm infestation both the mental and physical reactions are very slow. Little interest is taken in surroundings. The child does not run and play with other children. He is dull, apathetic, unable to concentrate upon his school problems, and apparently does not hear when spoken to, though slowly comprehending and gradually fulfilling appointed tasks. One has the impression that the child is living in another, entirely separate world, and is only remotely in contact with the everyday world about him. Every teacher of a southern rural school in the hookworm area has noted the immediate and almost startling improvement in mental alertness that occurs following treatment of these heavily infested children.

The degree of severity and the extent of the distribution of hookworm disease in the United States is of broad general interest to educators as well as public health workers for there is a large area in the group of Southern States which possesses the potential factors of climate, soil, sanitary customs etc., which are favorable for hookworm infestation. The group of the population which is chiefly affected by the disease in the Southern States is the rural white children of school age. This is the very group that by force of economic necessity is already handicapped by more limited educational opportunities than the town or city children, and it is only just that these rural children should be relieved of every possible handicap that has any retarding influence upon their mental and physical development in this most vital formative period of their lives.

Stiles (1915, *a*) who first discovered hookworm infestation in the United States strongly emphasized the physical and mental retardation in children due to hookworms. The disease became popularly known as "lazy man's disease," because of the physical langour and mental apathy of the unfortunate victims. Doctor Stiles (1915, *b*) made

[1] This study was conducted under the joint auspices of the International Health Board Hookworm Research Laboratory at Andalusia, Alabama, and the Alabama Department of Education.

memory tests of urban school children in the south and noted that *light infestation* with hookworms (Necator) did not seem to have any material effect in inhibiting the memory span of the infested children.

Strong, in collaboration with Stiles (1916) also studied the effects of hookworm disease upon mental development of children. The authors showed that hookworm disease interferes radically with mental development and felt that the longer a child had had the disease, the more he would lose mentally and the less rapid the gain after treatment.

Waite and Neilson (1919) in Australia, made mental tests on 340 children infested with hookworm, by the Binet-Simon scale and the Porteus Mazes. They attempted to divide the children into rough intensity groups, but at that time there was no satisfactory method of estimating hookworm intensity. They concluded that hookworm disease produced measurable mental sluggishness and that the retardation in mental development was in proportion to the massiveness of the infestation.

Washburn (1925) in Jamaica compared his hookworm records with the school records made by the teachers in groups of school children. He found that the daily records of the child in school presented a high correlation with the degree to which the child was suffering with hookworm disease. He estimated that a large percentage (possibly 75 per cent) of the mental retardation in the groups of school children studied was the result of hookworm infestation.

The purpose of the present study was to determine the degree of mental retardation of school children which could be attributed to hookworm infestation. Stoll has devised a method of measuring the number of hookworms harbored in an individual by a technique of ova counts of the stool. Smillie and Augustine, (1925) using this method of estimating the number of hookworms harbored by an individual, showed that the degree of physical retardation in school children is in proportion to the intensity of hookworm infestation. These authors classified the children into six hookworm intensity groups as follows:

Group I	No hookworms
Group II	Very light infestation (1–25 hookworms)
Group III	Light infestation (26–100 hookworms)
Group IV	Moderate infestation (101–500 hookworms)
Group V	Heavy infestation (501–1000 hookworms)
Group VI	Very heavy infestations (1001–3000 hookworms)

They measured 2000 school children in South Alabama by standard methods and found that the children who were heavily infested with

hookworms showed marked physical retardation. The lightly infested groups, however, showed no variation from normal cases. These cases of light hookworm infestation have been classified as carriers and not true cases of hookworm disease. The plan of procedure in the present study was similar to that of Smillie and Augustine in that the children were measured by standard mental tests and the intensity of their hookworm infestation was also determined. It was hoped that by correlation of these data, we might be able to measure the degree of mental retardation which accompanied varying intensities of hookworm infestation.

Methods of Procedure

An area in Covington County, South Alabama, was chosen in which there was a fairly heavy hookworm infestation. Three rural schools were selected. All children having obvious mental or physical defects were excluded from the series. The children were all whites of American parentage, living on small farms. The economic and social status of the children was closely comparable. The only outstanding variable was the variation in intensity of hookworm infestation.

The stools of the children were first examined to determine the intensity of their hookworm infestation by the Stoll ova count method. These observations were made by the hookworm research laboratory at Andalusia, Alabama. Physical measurements of the children were also made, the data being included in a separate study. The mental tests were made entirely independently of the physical tests and only after all data had been collected and compiled, was any attempt made to correlate the findings.

The mental tests were made as follows: The children in grades three to seven inclusive in the three rural schools mentioned above, were given three group intelligence tests, namely, the Otis Group Intelligence Scale and two forms of the National Intelligence Tests. In order that as nearly uniform conditions as possible might exist for each test, they were all given by the same person during one week. The scores for each test were converted into comparable scores by reducing them all to intelligence quotients according to calculations prescribed by the authors of the tests. There was almost uniform correlation between the scores for the Otis scale and the two scores for the National tests, the latter two tests being about 20 to 25 per cent lower than the Otis score in nearly every case. The intelligence

quotients used in this study are those obtained from the Otis scale alone. The names of children not taking the Otis scale and at least one form of the National test were eliminated. The final list contained the names of 118 children with their respective intelligence quotients as measured by the Otis scale, each one of these scores having been verified by one or two scores on the National tests.

Analysis of Results

The data were first analyzed in relation to Intelligence Quotient by dividing the children arbitrarily into groups in which every 10 points increase in IQ was considered as a separate group. We then determined the average number of hookworms harbored per case in each IQ group. The results of this analysis are given in Table I.

Table I shows that there is an inverse ratio between the average number of hookworms harbored and the Intelligence Quotient. The children having the highest intelligence averaged only 30 hookworms each, while the lowest intelligence group averaged 281 worms each.

A better way to present these results, perhaps, would be to divide children into hookworm intensity groups and then determine the Intelligence Quotient for each group. We have classified the children in hookworm intensity groups according to the plan of Smillie and Augustine, except that all children harboring 500 worms or more, are placed in one group and called heavy infestations.

Table I.—Intelligence Quotient in Relation to Estimated Intensity of Hookworm Infestation

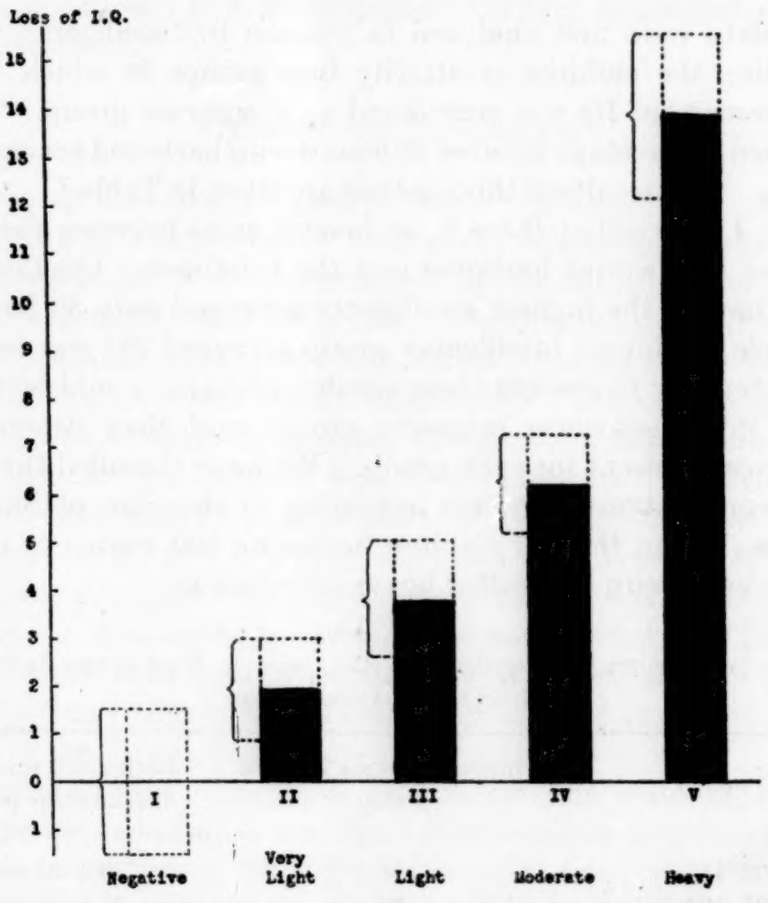
IQ group	Number of cases in each group	Estimated number of hookworms per case
105–115	4	30 hookworms
95–104	18	47 hookworms
85–94	45	67.8 hookworms
75–48	36	172 hookworms
65–74	15	281 hookworms

We then calculated the mean IQ with probable error for each intensity group, together with the standard deviation and its probable error and the coefficient of variation. The results of this analysis are given in Table II.

The results of the analysis given in Table II show that the heavily infested children—Group V—have a much lower Intelligence Quotient than the normal group, *i.e.*, children having no hookworms at all. The moderate infestation group (101–500 hookworms) also have a

Chart I.

Mental Retardation in Varying Intensity Groups of Hookworm Infestation.



Section of bar covered by bracket represents probable error.

definite mental retardation though not so marked as the retardation of the heavily infested children. The children with light infestations (26–100 hookworms) show a very slight retardation, while the group of very light infestations (harboring 1–25 hookworms) have no mental retardation at all, for they show approximately the same IQ as the normal children (within the limits of probable error of the calculation).

The degree of mental retardation caused by hookworm infestation in the various intensity groups is shown graphically in Chart 1.

It is obvious that not all heavily infested children were found to be stupid nor was every child without hookworms found to be exceptionally intelligent. There were two children in the group of 17 negative cases that had an IQ below 74, and none of these normal children scored higher than 104.

TABLE II.—THE INTELLIGENCE QUOTIENT OF VARIOUS HOOKWORM INTENSITY GROUPS

Intensity group	Number of cases	Mean IQ	Standard deviation	Coefficient variation
Negative	17	90.2 ±1.53	9.496 ±1.10	10.5
Very light infestation 1–25 worms	40 ..	88.3 ±1.06	9.96 ±.75	11.2
Light infestation 26–100 worms	27 ..	86.4 ±1.2	9.54 ±.875	11.0
Moderate infestation 101–500 worms	23 ..	84.1 ±1.07	7.68 ±.76	9.13
Heavy infestation 501–2000 worms	10 ..	76.3 ±1.737	8.136 ±1.22	10.7

On the other hand, one child harboring over 500 hookworms had an IQ of well above the average,—97. She was the only one of the heavily infested group, however, with an IQ greater than 84. The distribution of the cases of the various intensity groups in relation to each IQ group is given in Table III.

Table III shows clearly that though there are certain children who may score an Intelligence Quotient as high or higher than the average in spite of heavy hookworm infestation, yet the general rule is that the heavier the hookworm infestation, the lower the Intelligence Quotient. Only the lightly infested group of cases shows no deviation from the normal.

TABLE III.—NUMBER OF CASES FROM VARIOUS HOOKWORM INTENSITY GROUPS IN EACH IQ GROUP

IQ group	Intensity of Hookworm Infestation					
	Negative cases	Very light	Light	Moderate	Severe	Total cases
105–115	0	3	1	0	0	4
95–104	8	6	2	1	1	18
85–94	5	14	16	10	0	45
75–84	2	14	4	10	6	36
65–74	2	3	4	2	3	14
Total cases.....	17	40	27	23	10	117

DISCUSSION

The results of this investigation closely parallel those obtained by Smillie and Augustine in their analysis of the physical retardation in school children which is produced by varying intensities of hookworm infestation. These authors showed that the degree of retardation in growth in weight and height, and the lowering in hemoglobin was in proportion to the intensity of hookworm infestation. We have shown that the same rule holds when we measure the mental retardation in school children which is produced by hookworm infestation. We were already fully aware of the fact that hookworms produced mental retardation in growing children, but this study has enabled us to measure the actual degree of injury produced by a given number of hookworms. Another important point brought out by this analysis is that very light hookworm infestations produce no measurable harm to growing children. These children are hookworm carriers and not true cases of hookworm disease.

This fact has an important practical bearing in the control of the disease. It is a comparatively simple matter to reduce all cases of hookworm disease to the point of economic cure by one, or at most two, standard treatments. Sometimes two or three additional treatments are required to remove the few remaining worms which are causing the children no measurable harm. It is obviously unjust to submit a child to the discomfort of a vigorous hookworm treatment when there is no hope of a compensatory advantage to be gained in increased mental and physical vigor.

The best plan of procedure therefore in an area where school children are infested with hookworms is to examine the school children

at yearly intervals, not in relation to incidence of hookworm infestation, but hookworm intensity. Children with very light infestations may be considered as carriers and need not be given hookworm treatment, for the small number of hookworms harbored are causing them no measurable mental or physical injury. The children that suffer from actual hookworm disease, however, should be given treatment at once with the fullest assurance that a rapid, and in the severe cases, a striking improvement in mental and physical vigor will occur in these children.

Conclusions

1. Rural white school children with heavy hookworm infestation—500 or more *Necator americanus*—show a marked degree of mental retardation.

2. School children with a moderate intensity of hookworm infestation (100–500 hookworms) have a slight degree of mental retardation.

3. Very light intensities of hookworm infestation (1–25 hookworms) show no variation in their Intelligence Quotient from normal children. These cases of light hookworm infestation are carriers and not true cases of hookworm disease.

Literature Cited

Smillie, W. G. and Augustine, D. L.: The Effect of Varying Intensities of Hookworm Infestation upon the Physical Condition of School Children. *American Journal of Diseases of Children*, Vol. XXXI, pp. 151–168.

Stiles, C. W.: (*a*) The School Grades Attained by 2166 White School Children in the City of *X*, Classified by Age, Sanitation and Intestinal Parasites., U. S. P. H. Reports, Vol. XXX, 1915, 2060,

(*b*) Memory Tests of School children., *ibid.*, Vol. XXX, 3738, 1915.

Strong, E. K. and Stiles, C. W.: Effects of Hookworm Disease on the Mental and Physical Development of Children., International Health Commission No. 3, 1916.

Waite, J. H. and Neilson, I. L.: Effects of Hookworm Disease on Development of North Queensland Children., *Jour. Am. Med. Assn.*, Vol. LXXIII, pp. 1877–1879.

Washburn, B. E.: Report of Jamaica Hookworm Commission for 1924. Gov. Printing Office, Kingston, Jamaica. 1925.

AN INVESTIGATION OF THREE PLANS FOR SELECTING THE STUDENTS TO BE ADMITTED TO COLLEGE

ESTHER CRANE[1]
Goucher College

Many colleges today are interested in comparing the relative merits of entrance examinations and of psychological tests as methods for selecting the students to be admitted to college, because they are so overcrowded that they are unable to admit all of the applicants who fulfill their entrance requirements. Such colleges usually endeavor to select from their applicants as many as possible who will do excellent work, and as few as possible who will do poor work in college. Therefore, a most practical question for any college which admits students only through the regular channels of entrance examinations and which is unable, because of its lack of accommodations, to admit all of the students who pass its entrance examinations is this: Which of these three alternative plans will succeed in selecting the largest proportion of students who will do good work in college and the smallest proportion of students who will do poor work—(1) competitive entrance examinations which exclude those students who make the lowest records in entrance examinations; (2) competitive psychological tests which exclude those students who have the lowest psychological scores no matter what their records in their entrance examinations may be; (3) a combination of these methods which excludes only those students who have poor records in entrance examinations and also poor records in the psychological test used. The following discussion will attempt to answer this question in terms of an investigation which has been carried on for four years at Bryn Mawr College.[2]

---

[1] The writer wishes to acknowledge her indebtedness to Miss Edith Orlady, Secretary and Registrar of Bryn Mawr College, for her cooperation and courtesy in furnishing data as to entrance examination averages and records of grades in college, and also to acknowledge her indebtedness to the following students for help in making the correlations and tables used in this paper: Miss Katherine Elizabeth McBride, Miss Charlotte Fisk, Miss Margaret Elizabeth Pillsbury, Miss Janetta Wright Schoonover, Miss Doris Stearns.

[2] The investigation was begun by Dr. Ada H. Arlitt and a comparison of the Thurstone test scores for the classes of 1924 and 1925 with their records in entrance examinations and in freshmen college work, and for 1925 also with their records for sophomore college work, is reported by Dr. Arlitt in the *Journal of Applied Psychology*, Vol. VII, December, 1923, pp. 330–338.

In October 1920 the class of 1924 at Bryn Mawr College was given the Thurstone Test IV, 1919 edition; in October 1921 the class of 1925 was given the Thurstone test, 1920 edition; in October 1922 the class of 1926 was given the Thurstone test, 1922 edition; and in October 1923 the class of 1927 was given the Thurstone test, 1922 edition. All of these classes were admitted to college through the regular channels of entrance examinations and their scores in the Thurstone tests had no influence on their admission to college or on their marks in those subjects which they studied in college.[1] The results of the psychological tests and of the college entrance examinations have been compared with each other, and each has been compared with the grades received in college; in the case of the class of 1924 with their college marks for four years, in the case of the class of 1925 with their college marks for three years, in the case of the class of 1926 with their college marks for two years, and in the case of the class of 1927 with their college marks for one year. Grades in Bryn Mawr College are given in terms of High Credit, Credit, Merit, Pass, and Fail. For each hour of High Credit work the student is given three honor points; for each hour of Credit work she is given two honor points; and for each hour of Merit work she is given one honor point. Pass and Fail are alike therefore, in counting nothing at all toward the honor point score. The comparisons in this paper are all in terms of this honor point score.

Beginning with a consideration of the first plan we may investigate in the case of each of the four classes whose records are available what would have been the results of making the entrance examinations competitive and excluding the lowest 10 per cent of the students. Table I presents the college records of the 10 per cent who would thus have been eliminated. The first section of this table shows us for all four classes under consideration how many of the students who would have been excluded in this case were students who fell below the twenty-fifth percentile of their class in their first year's college work, how many were included between the twenty-fifth and fiftieth percentile, how many between the fiftieth and seventy-fifth, how many of them ranked above the seventy-fifth percentile, and how many of them withdrew from college before the end of the first year. The second section of the table gives this same distribution for three of these

[1] The individual scores in the psychological tests have never been made known to any members of the faculty or to any of the students, so that it is not possible that these psychological scores in any way affected the judgment of the faculty in marking the students or the decisions of the students in choosing their courses.

classes, 1924, 1925 and 1926 at the end of two years' work, the third section shows the distribution for 1924 and 1925 at the end of three years' college work, and the last section of the table gives the distribution for 1924, the only class for which we have a complete record for all four years of college. These figures will help us to determine whether it is true that low grades in entrance examinations do generally predict poor work in college, and whether this first plan of making the entrance examinations competitive would exclude more poor students than good ones.

TABLE I.—THE LOWEST 10 PER CENT OF THE CLASSES OF 1924, 1925, 1926, AND 1927 WHEN ARRANGED ACCORDING TO THEIR MARKS IN ENTRANCE EXAMINATIONS

	Lowest quarter in honor points	Second quarter in honor points	Third quarter in honor points	Highest quarter in honor points	Withdrawn	Total number
	At End of Freshman Year					
Lowest 10 per cent in 1924 in entrance examinations 62.9–65.15	3	5	2	0	0	10
Lowest 10 per cent in 1925 in entrance examinations 57.7–62.15	7	4	1	0	0	12
Lowest 10 per cent in 1926 in entrance examinations 58.1–64	6	2	3	0	1	12
Lowest 10 per cent in 1927 in entrance examinations 62.2–65.6	7	1	1	1	2	12
Total of lowest 10 per cent	23	12	7	1	3	46
	At End of Sophomore Year					
Lowest 10 per cent in 1924 in entrance examinations 62.9–65.15	4	3	0	1	2	10
Lowest 10 per cent in 1925 in entrance examinations 57.7–62.15	3	2	1	0	6	12
Lowest 10 per cent in 1926 in entrance examinations 58.1–64	4	2	3	0	3	12
Total of lowest 10 per cent	11	7	4	1	11	34
	At End of Junior Year					
Lowest 10 per cent in 1924 in entrance examinations 62.9–65.15	3	2	0	1	4	10
Lowest 10 per cent in 1925 in entrance examinations 57.7–62.15	1	3	2	0	6	12
Total of lowest 10 per cent	4	5	2	1	10	22
	At End of Senior Year					
Lowest 10 per cent in 1924 in entrance examinations 62.9–65.15	3	2	1	0	4	10

Let us examine the first section of Table I which gives for the four classes considered in this report an estimate of the quality of work done in their freshman year by the 10 per cent who would have been excluded by the first plan, that of making the entrance examinations competitive. Out of the 46 students who would thus have been excluded 35 did work which was below average in their freshman year, 3 withdrew before the end of their first year, and only 8 did work above the average. Therefore this first plan of making the entrance examinations competitive would have excluded almost 5 students who did work below the average or withdrew from college before the end of the first year for every 1 student who did work above the average in the first year. It is also significant that half of the students who would have been excluded by this plan were in the lowest quarter of their class in freshman year, whereas there was only 1 out of the 46 who was in the highest quarter of the class. When we examine the records of the three classes whose marks we have for two years of college work, and likewise the records of the two classes whose marks we have for three years, we find that this plan would have eliminated 6 students who did work below the average or withdrew from college for every 1 student who did work above the average. There is only one class whose record we have for the whole college course, so our evidence on that point is slight, but in this class only 1 student whose marks were above the average would have been eliminated and 9 whose marks were below the average or who withdrew before the end of the four years work. Therefore we may judge that this first plan for determining admission to college would certainly have excluded from the four classes considered many more students who fell below the average in their college work than students who did work above the average. However, the proof that this first plan of making the entrance examinations competitive would have excluded 5 or 6 poorer students for every 1 better student shows only that this is one good principle for a college to use in selecting students, it does not show that it is the best of the three plans discussed.

Let us now examine the results which would have followed if the college, after giving the Thurstone test to those four classes whose records we are considering, had excluded the 10 per cent whose marks were lowest in this psychological test. Table II shows these results arranged according to the same plan which was followed in Table I.

TABLE II.—THE LOWEST 10 PER CENT OF THE CLASSES OF 1924, 1925, 1926, AND 1927 WHEN ARRANGED ACCORDING TO THEIR MARKS IN THE THURSTONE TEST

	Lowest quarter in honor points	Second quarter in honor points	Third quarter in honor points	Highest quarter in honor points	Withdrawn	Total number
	At End of Freshman Year					
Lowest 10 per cent in 1924 in Thurstone test 66–87	5	3	1	1	..	10
Lowest 10 per cent in 1925 in Thurstone test 60–84	5	4	1	2	..	12
Lowest 10 per cent in 1926 in Thurstone test 12–20	4	6	2	0	..	12
Lowest 10 per cent in 1927 in Thurstone test 17–121	3	6	3	0	..	12
Total of lowest 10 per cent	17	19	7	3	0	46
	At End of Sophomore Year					
Lowest 10 per cent in 1924 in Thurstone test 66–87	4	1	1	2	2	10
Lowest 10 per cent in 1925 in Thurstone test 60–84	5	1	2	1	3	12
Lowest 10 per cent in 1926 in Thurstone test 12–20	2	3	2	0	5	12
Total of lowest 10 per cent	11	5	5	3	10	34
	At End of Junior Year					
Lowest 10 per cent in 1924 in Thurstone test 66–87	4	1	1	2	2	10
Lowest 10 per cent in 1925 in Thurstone test 60–84	5	2	2	0	3	12
Total of lowest 10 per cent	9	3	3	2	5	22
	At End of Senior Year					
Lowest 10 per cent in 1924 in Thurstone test 66–87	3	3	1	1	2	10

The comparison of this table with Table I convinces us that although this second plan of admission to college would have eliminated more poor students than good ones it would not have discriminated between them as successfully as the first plan of making the entrance examinations competitive. An examination of the records of all four classes for their freshman year shows that this plan of making the Thurstone test competitive would have excluded 10 students who were above the average and 36 students who were below the average

in their freshman work, which amounts to the exclusion of 1 more successful student for every 3½ less successful students. Inasmuch as the first plan, that of making the entrance examinations competitive, would have excluded only 1 of the more satisfactory students for every 5 less satisfactory ones, and inasmuch as this second plan would have eliminated 3 students who stood in the very highest quarter of their class freshman year whereas the first plan would have eliminated only 1 of these students in the upper quarter, it seems obvious that the plan of making the Thurstone test competitive is distinctly inferior to the plan of making the entrance examinations competitive so far as these results for the freshman year are concerned. The consideration of the classes for which we have longer records tends to indicate a still greater inferiority of the second plan as compared with the first. From the three classes whose records we have for two years, and likewise from the two classes whose records we have for three years, this plan would have excluded 1 student who did work above the average for every 3 students who did work below the average or who withdrew from college before the end of the year considered. From the one class whose records we have for all four years there would have been 1 student whose work was above the average excluded for every 4 whose work was below the average or who did not finish the four years' course. When we remember that the first plan would have eliminated 6 less satisfactory students for every 1 more satisfactory student when measured by two or three years' work, and 9 less satisfactory for every 1 more satisfactory student when measured by four years' college work, we realize that this second plan would have been decidedly inferior to the first in choosing from the four classes considered those students who did the best work in Bryn Mawr College. However, this plan cannot be considered entirely unsuccessful, since it would eliminate 3 poorer students for every 1 better student.

Lastly let us examine the third plan for selecting students for admission to college, the one which combines the evidence of the entrance examinations and the psychological tests and excludes only those students who fall in the lowest quartile of both. This plan has a certain theoretical justification in that it excludes only those students who rank lowest in their group in both of these independent measures which have been shown to be somewhat indicative of college success. The facts presented in Table III tend to bear out this theory. This table presents the records of those students who fell in the lowest quartile of their class both when it was divided according to entrance

examinations and also when it was divided according to scores in the Thurstone test, that is, those students who show no particular ability in either line.

TABLE III.—STUDENTS FROM THE CLASSES OF 1924, 1925, 1926, 1927 WHO FELL IN THE LOWEST QUARTILE IN BOTH ENTRANCE EXAMINATION AND THURSTONE INTELLIGENCE TEST

	Lowest quarter in honor points	Second quarter in honor points	Third quarter in honor points	Highest quarter in honor points	Withdrawn	Total number
	At End of Freshman Year					
Lowest quarter in 1924 in both entrance examinations and Thurstone test.....	3	2	2	0	0	7
Lowest quarter in 1925 in both entrance examinations and Thurstone test.....	6	4	0	0	0	10
Lowest quarter in 1926 in both entrance examinations and Thurstone test.....	7	2	3	0	0	12
Lowest quarter in 1927 in both entrance examinations and Thurstone test.....	3	5	2	0	1	11
Total number in lowest quarter of both entrance examinations and Thurstone test.....	19	13	7	0	1	40
	At End of Sophomore Year					
Lowest quarter in 1924 in both entrance examinations and Thurstone test.....	3	1	1	1	1	7
Lowest quarter in 1925 in both entrance examinations and Thurstone test.....	5	1	0	0	4	10
Lowest quarter in 1926 in both entrance examinations and Thurstone test.....	4	2	1	0	5	12
Total number in lowest quarter in both entrance examinations and Thurstone test.....	12	4	2	1	10	29
	At End of Junior Year					
Lowest quarter in 1924 in both entrance examinations and Thurstone test.....	2	1	0	1	3	7
Lowest quarter in 1925 in both entrance examinations and Thurstone test.....	4	1	0	0	5	10
Total number in lowest quarters of both entrance examinations and Thurstone test.....	6	2	0	1	8	17
	At End of Senior Year					
Lowest quarter in 1924 in both entrance examinations and Thurstone test.....	2	1	1	0	3	7

As can be seen from Table III this third plan would exclude almost the same number of students as the first and second plans, so that a direct comparison of Table III with Tables I and II will show definitely which of the three plans under consideration would have succeeded in excluding the largest proportion of poorer students and the smallest proportion of better students from the four Bryn Mawr classes whose records are available. In the first section of Table III we find that the records for all four classes for freshman year show 32 students whose work was below the average and 1 who withdrew before the end of the first year as compared with 7 students who did work above the average. This amounts to the exclusion of almost 5 students who did work below the average or withdrew from college before the end of the first year for every 1 who did work above the average, which is exactly the same proportion that would have been obtained by the use of the first plan, and a very much better proportion than that which would have been obtained by the use of the second plan. So far as the results of one year's work are concerned, the only slight claim to superiority over the first plan which the third plan possesses is that it would have excluded no student who stood in the highest quarter of her class during her freshman year, whereas the first plan would have excluded 1 such student. But when we take into consideration the results of two years' work in college we find that the third plan would have eliminated 9 less satisfactory students for every 1 more satisfactory student, whereas we saw that the first plan would have excluded only 6 less satisfactory for every 1 more satisfactory student. This superiority of the third plan is even more striking when the results for three years' work are considered, because there is only 1 of these students who would have been excluded who has done work above the average, as compared with 16 who did work below the average or withdrew from college before the end of the three years. The first plan cannot show any results as good as that, in fact the group excluded by the first plan contained only 6 less satisfactory students for every 1 more satisfactory student when three years' work was concerned. The only place where the third plan has no claim at all to superiority over the first plan is in the one class of 1924 for which we have records for all four years. From this class there would have been only 7 students excluded by the third plan, and 1 of these did work above the average when all four years in college are considered; whereas the first plan would have excluded 10 students from that same class of whom only 1, and that this same student, did work above the average in college. With this

exception all the evidence we have points to the superiority of the third plan over the first plan of determining admission to college. It has not seemed necessary to show in detail the superiority of the third plan to the second, since it has already been shown that the second plan is inferior to the first.

Conclusions

Three plans of determining admission to college in cases where all of the applicants cannot be admitted have been investigated by examining the marks actually obtained in college by the groups of students who would have been excluded from four classes of Bryn Mawr students under the operation of each of these three plans. If we measure the success of these plans by their ability to exclude a small proportion of students whose work was above the average and a large proportion of students whose work was below the average or who withdrew from college before the end of their college course, we find:

1. That the first plan, that of making the entrance examinations competitive and excluding the 10 per cent who made the lowest marks in entrance examinations, was superior to the second plan and inferior to the third plan in all the classes whose records have been obtained.

2. That the second plan, that of making the psychological test competitive and excluding the 10 per cent who made the lowest marks in the psychological test was inferior to the other two plans in the four classes which have seen considered.

3. That the third plan, that of combining the results of the two measures used and excluding those students whose marks not only fell in the lowest quarter of entrance examinations but also in the lowest quarter of psychological tests, is superior to the other two plans in all the cases studied.

It must be admitted that not one of these three plans succeeds in excluding only those students whose work is below the average. Not only would they all have excluded some students from the classes discussed whose work proved to be above the average, but each one of them would have excluded a few students from some of the classes investigated whose marks placed them in the highest quarter of the class. Therefore these three plans are far from perfect, even according to the criterion used here in judging them. Nevertheless any one of them might prove extremely useful to a college which is unable to admit all applicants, and such a study of the records made in college by several classes would easily show which of the three plans would be most useful for any given college.

A METHOD FOR EVALUATING A PSYCHOGRAPH FOR VOCATIONAL GUIDANCE[1]

J. B. MINER

The University of Kentucky

Very little progress has been made in the development of scientific procedures for vocational guidance. The problem is essentially different from that for the selection of employes, toward which considerable progress has been made. If the plans for research on vocational adjustment should be controlled by those who were interested only in finding employes who could serve them, the program would naturally neglect the vocational guidance phase of the problem which seeks a permanently satisfying occupation for the individual. In general an employer desires to know whether a prospect can do the work in his particular establishment. He is necessarily interested in his own business, not in the question whether an individual would be happier under another employer or in a different line of work. He would be glad to see prospects who are not adapted to his business placed in the vocations for which they are best adapted. It is only a short step from this to the public spirited and far-seeing employer who recognizes that all are best served when each individual is in the work which is most permanently satisfying to himself.

Vocational guidance research will advance most rapidly when such public-spirited employers, universities, and independent educational organizations can be brought together for its program. A case in point is the cooperative study of the vocational adjustment of college students organized through a committee of the National Research Council. The research is to be planned by an advisory council, with Dean H. E. Hawkes of Columbia as chairman. The council includes representatives from each of 14 universities and from the American Council on Education. It has been assured financial support from several research organizations and progressive business and industrial concerns. In the analysis of its problem it definitely recognized two phases: "Development of more objective and effective methods of measuring the qualifications of men" and "the definition and analysis of the requirements of various professions." It expects to begin with the professions of medicine and business. If a number of important

[1] Read before the 1925 meeting of the Southern Society for Philosophy and Psychology.

and slightly correlated human traits are properly measured they can later be weighted so as to predict their importance in other occupations studied. In this way the measurements will become available for a real guidance program. It is necessary to begin with a very few vocations, but at the same time to guard against sending graduates into business and medicine when they might be better adapted by interest, aptitude, and training for other fields of work. To make a start toward guidance it is essential that special personal characteristics of widely varying importance for various situations shall be measured, to supplement tests of intelligence which measure factors commonly valuable in all intellectual situations.

Some years ago, on account of the obvious importance of special interests for the problem of vocational adjustment, the writer experimented with a new method of recording interests in relation to work conditions. At the meeting of the American Psychological Association in 1920 in Chicago, he read a paper on "An Interest Psychograph" which described the plan for securing records which is utilized in this paper. This self-analysis record by means of paired interest contrasts formed part of a blank for the Analysis of Work Interests.[1] It was later included in the personal history blanks which the students filled out on entrance at the Carnegie Institute of Technology and at the University of Kentucky. A method for recording work interests indirectly has been developed by Moore, Ream and Freyd.[2] It consists of recording likes, dislikes and indifference toward certain occupations, for example, actor, architect, artist, astronomer, etc. Whether it is a more useful method of self-analysis for vocational interests than the paired contracts has not been tested.

The method of paired interest contrasts was tried in a preliminary way to distinguish small groups of students in different vocational courses at the Carnegie Institute of Technology. In the paper above referred to it was shown that it would separate women taking the secretarial course from those taking home economics, or men in the course for machine construction and operation from those in the course for electrical construction and equipment. By means of critical scores it is possible to distinguish groups of individuals roughly.

---

[1] Miner, J. B.: An Aid to the Analysis of Vocational Interests. *Journal of Educational Research*, April, 1922.

[2] Freyd, Max.: A Method for the Study of Vocational Interests. *Journal of Applied Psychology*, Vol. VI, 1922, pp. 243–254.

Ways of accomplishing this have been described by Thurstone, Bills,[1] and Freyd. Without using partial regression methods which would become exceedingly cumbersome for a long series of scores, there seemed to be no convenient method by which the various tests of a series could be conveniently weighted so that the total profile of an individual could be compared to that of a number of typical groups. This is an essential problem of method in a vocational guidance program. It is the aim of the present paper to suggest a rather simple method for meeting this more general problem. It can be conveniently used at least for the preliminary sifting of the series of tests to select those few which are most distinctive for each vocation. It has the merit of also permitting an error of prediction for the indices of a particular person to be empirically calculated when the series of tests is long enough to be sub-divided.

The data which are here utilized for illustrating the method for evaluating a psychograph consist of the records of some 800 Kentucky students. One hundred were taken alphabetically from each of the four groups of students who indicated that their first choice for their own life work was teaching, engineering, law, or medicine. The plan suggested here also requires a random group to which each of these pseudo-occupational groups is compared. The random group used consisted of 570 cases taken alphabetically from all entering students regardless of whether they chose these four occupations or others. This proved to be sufficiently at random to bring out differences in the professions. The pseudo-occupational group of 100 prospective teachers included 80 women. The other occupational groups were made up of men. It is obvious that the use of data from such pseudo-occupational groups is jutified only for the purpose of trying out a method to see whether it promises to be sufficiently useful to justify an attempt to secure data with groups of people who have been long established in their vocations. If the method of evaluating the data seems feasible and funds were available it would be a comparatively simple matter to obtain similar data from people actually engaged in these and other professions.

One other statement as to the data used for this study should be made. A hundred students who had filled out these interest contrast records were asked to fill out the same blank three months later. There

---

[1] Bills, M. A.: Methods for the Selection of Comptometer Operators and Stenographers. *Journal of Applied Psychology*, Vol. V, 1921, pp. 275–283.

were ten interest contrasts out of the original 37 for which the reliability coefficients fell below .69. The coefficients were calculated on the basis of the percentage changing from one of the pair to the contrasting interest, disregarding those who changed only between marking neither and marking one of the pair on the assumption that these completed but did not shift their records. All except one of these less reliable contrasts were omitted. That contrast having to do with working by one's self or with others, was retained although its reliability coefficient was only .51, since it was one of the five which most distinguished the premedic group.

The reliability coefficients indicate merely the fact that the same individuals were able a few months later to size their interests up in about the same way. The question might be raised whether a self-estimate of any kind is really a sufficiently reliable criterion of interests to justify any measurement procedure. If suitable objective tests were available, they might be expected to provide better data which could be evaluated by the same method. It is not at present clear, however, that any objective measurement of interests which has yet been proposed is as closely related to actual work satisfactions or even occupational success as these apparently cruder but more vital self-estimates in which an individual is called upon to judge his own *potentiality*, such potentiality may be mainly a native trend which no objective test can reach so well. At any rate this form of record is worth testing provided that it can be put in quantitative form. Even without measurement, these paired interest contrasts are useful in connection with the study of an individual's vocational choice. The checking of these contrasts may disclose inconsistencies with the conditions of work which may be expected in the vocations considered. If the choice lies within two or three occupations it may assist in completing the decision. They may also be used to disclose very unusual interests when an individual checks a pair differently from the great majority. The usefulness of the record of interests will, however, be materially increased if a satisfactory method of evaluating the total profile of these contrasts can be devised.

The plan submitted makes the assumption that any one of these pairs of work interests would be distributed normally in a random group of adults. If this is accepted as a working hypothesis, then it is possible to convert the percentage choosing, for example, indoor work as a contrasted with outdoor work into the average standard deviation

value of such a percentage. This can be readily done by Thorndike's table[1] for transmuting percentages.[2]

Using the random group for this common valuation of a work interest, the largest percentage may always be placed toward the plus end of the distribution and the contrasting choice toward the minus end of the distribution. Those who fail to check either one of the pair are supposed not to be able to distinguish a difference in themselves in respect to this pair of interests. The percentage checking neither one of the pairs may thus also be transmuted into a sigma value by giving it the position between the two extreme groups. For example, the percentages found for the random college group were

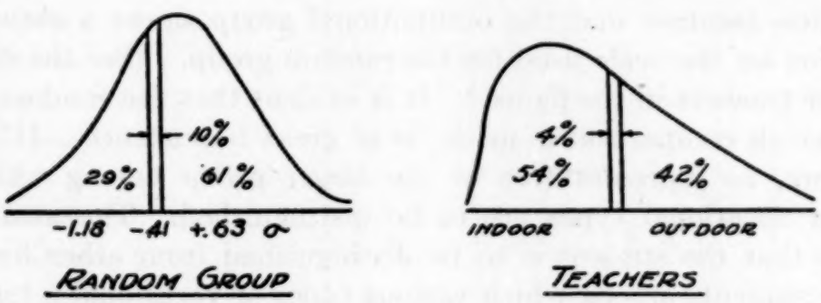


FIG. 1.

indoor 29 per cent, neither 10 per cent, outdoor 61 per cent. Transmuted into sigma values, the average sigma value for those checking *indoor* becomes −1.18 sigma; *neither* −.41 sigma; *outdoor* +.63 sigma.

The next problem is to obtain a valuation for the same interest contrast on the part of an occupational group. The shift in the percentage of the particular occupation compared with the random group having the same interest is regarded as distinguishing that occupation. The value of a certain type of interest, say "indoor work," for an occupation was determined by taking the sigma value for that interest in the random group, in this case, −1.18; and then multiplying it by the difference in percentage of the occupational group which checked it. For example, in the teacher's group, 54 per cent checked "indoor"

---

[1] Thorndike, E. L.: An Introduction to the Theory of Mental and Social Measurements, revised edition, Chap. VIII.

[2] In case the actual sigma values of the degrees of a trait are measured for a random group, as they would be for a test, these sigma values would take the place of the transmuted values of percentages. For example, an opposite test would give sigma values for each score.

as compared with 29 per cent for the random group. This difference of 25 was multiplied into 1.18 sigma giving the value 29.50 sigma. Dropping the decimal point and figures beyond it as unreliable we have the nearest value 30 which is the value stated for "indoor" under teaching on the evaluation table. All differences in which the occupation has a larger percentage with a particular interest than does the random group are treated as plus values. This means that the checking of that interest is more of an indication of that occupation than of a random choice. In the same way when fewer in an occupation check a particular interest that difference is given a minus value indicating that a person having that interest is less like that occupational group than he is like the random group. It may be noted that this plan of valuation assumes that the occupational group shows a skewed distribution on the scale used for the random group. (See the distribution for teachers in the figure.) It is evident that the random group, with which comparison is made, is of great importance. It should, therefore, be representative of the larger group among which the various vocational types are to be distinguished. The assumption here is that the student is to be distinguished from other first year college students among which various types of vocational interest are represented.

With the table of values at hand it is a simple matter to evaluate the record of a particular individual for each of these four pseudo-professional groups and to state in numerical terms what is his total record on the 28 contrasts. For example, one student shows totals of −33 for teaching, +62 for engineering, +18 for law, and −23 for medicine. He has a margin of 44 for engineering over its closest rival, law. His first choice as indicated by him at the time the blank was filled out was engineering with a second choice to be a musician.

For convenience the use of the entire 28 contrasts may be simplified by using only the five largest and most characteristic differences from the random group for each profession. It may also be checked by the number among the 28 interests which the individual checks in the same direction as the occupation away from the random group, disregarding the weights given to each contrast. If two or three of these ways of treating the data agree, the indication of the favored vocation is, of course, more reliable.

To get an idea of how the method worked with the rather unsettled choices of these college students, the records were evaluated for 20 students whose first choice was in each of these four professions. Out

TABLE FOR EVALUATING THE PSYCHOGRAPH OF WORK INTERESTS

Notice the contrasts in working conditions listed below. See how they are related to your own interests and abilities. Whenever you can distinguish a difference in yourself, check that one in the pair under which you believe you would do the better work. Consider in which of the pair, after training, you would trust yourself more to meet the competition of others. Failure to check a pair will indicate that you can not distinguish a difference in yourself in that contrast.

	Teaching	Engineering	Law	Medicine
Indoor	30	−18	11	−2
Neither	−3	0	2	3
Outdoor	−12	10	−9	−3
Skilled hand-work	5	−5	28	1
Neither	1	1	−20	0
Skilled heavy work	−15	10	−29	−2
Slow movements	6	−17	1	1
Neither	2	1	1	4
Rapid movements	−6	8	−2	−7
Broad planning	5	7	4	−9
Neither	0	0	1	1
Attention to details	−10	−10	−10	3
Less responsibility	24	−7	−7	−13
Neither	15	1	−1	3
More responsibility	−12	2	4	3
Organizing people	11	−23	30	1
Neither	0	−1	0	1
Constructing things	−6	16	−12	3
Working with things	−6	27	−24	5
Neither	−1	2	2	1
Working with people	5	−18	10	−4
Working by yourself	−7	2	−18	18
Neither	0	2	1	1
Working with others	4	−6	10	−14
Operating appliances	−3	1	−1	−1
Neither	6	−7	1	−4
Repairing appliances	−8	15	2	11
Doing the same thing	7	−10	2	13
Neither	0	−2	3	2
Wide variety in work	−2	3	−2	−4
Thoughtful hand-work	8	1	−20	−16
Neither	−1	2	−3	0
Thoughtful head work	−3	−3	13	1
Methodical work	15	−14	−10	−8
Neither	0	−1	1	2
Meeting new conditions	−6	7	3	2
Regular time for work	2	−1	−4	−6
Neither	−1	−1	1	3
Irregular time for work	−6	6	11	13
Work requiring energy	−7	5	0	−3
Neither	1	−1	1	−1
Work requiring patience	8	−6	−6	7
Work requiring judgment	−7	−1	5	3
Neither	0	−1	0	3
Work requiring industry	16	4	−13	−13
Work requiring quickness	−3	1	14	0
Neither	1	3	1	1
Work requiring reliability	1	−3	−7	−1

TABLE FOR EVALUATING THE PSYCHOGRAPH OF WORK INTERESTS.—*Continued*

	Teaching	Engineering	Law	Medicine
Work requiring speed	1	−4	3	−6
Neither	0	3	−3	1
Work requiring judgment	−1	−1	1	2
Work in one town	26	−26	−2	0
Neither	0	0	0	0
Work requiring traveling	−21	−12	1	−1
Larger future success	0	4	2	3
Neither	2	−1	−1	−3
Smaller immediate success	−3	−13	−5	−8
Influencing people directly	−2	−4	10	3
Neither	0	1	0	1
Influencing people indirectly	2	10	−13	−8
Making things as directed	6	−12	−17	−9
Neither	−1	−1	0	2
Following your own methods	−2	9	7	1
Work requiring expression	6	−14	9	−5
Neither	1	0	1	2
Work requiring computation	−6	18	−17	−1
Work in an office	−28	4	17	25
Neither	0	0	2	0
Work requiring teaching	40	−10	−10	−19
Welfare work	6	−5	−10	4
Neither	0	0	0	0
Taking part in entertainments	−4	−1	4	7
Work requiring artistic taste	2	−6	−1	−21
Neither	0	1	2	0
Work requiring artistic creation	−16	−3	−22	27
Work requiring common sense	3	−1	−1	−8
Neither	0	−1	0	3
Work requiring specific information	−7	3	1	10
Increasing the amount of production	−1	−3	−10	27
Neither	0	0	3	−4
Improving quality of the product	1	1	1	−9
Greater pay with risk or discomfort	17	14	5	24
Neither	0	0	0	0
Less pay with more safety and comfort	17	−11	−5	−17

of this total of 80 students, 71 had a total plus record in agreement with his choice. In 52 cases this was also the highest plus record of any occupation, using the most complete criterion, that of the 28 contrasts as weighted in the table. When one considers that a chance distribution would have placed only 20 in their chosen professions, this result is 2.6 times out of a possible four times better than chance. This seems to be a simple way of understanding the success of the method for comparison with other methods used with the same sort of data. In other words, 65 per cent correctly placed is 2.6 times chance for four professions, while it would only be 1.3 times chance if it were predicting between only two professions. This result of 2.6 times chance may be taken as a minimum value for the success of the method.

For a maximum evaluation of the method other factors should be taken into consideration in view of the uncertainty of the data used. We do not know how well the students chose among these professions the one best suited to their own interests. Let the 52 out of 80 whose choices were confirmed be regarded as correct. There were 24 others who were shown by two or by three of the methods of treating the data, to be really more alike in interest to those of another profession than to the one chosen. These two groups would then include all except 4 of the 80 students' records. That is, 52 probably chose their favored profession and 24 others might reasonably be regarded as more correctly placed in the profession indicated by the psychograph. This would leave only 5 per cent of the cases not properly placed by the method. This would be the most favorable showing for the method. Comparing the two extreme estimates, it seems safe to say the method placed correctly between 65 per cent and 95 per cent of these 80 students among four possible professions. Using this plan for evaluation, the success of the method may be taken to be from 2.6 times to 3.8 times chance.[1]

For the prediction of the favored profession for a particular individual it is desirable also to know the error of his index for each of the professions. If the standard table of values for each work interest were calculated for representative professional and random college groups, instead of these pseudo-professional groups, we might then accept these standards as correct for the purpose of comparison and ask regarding a particular person which of these representative groups does his record most resemble. This would leave the question "what is the error of his index?" for each profession which would show if we took comparable samples of his interests. Since we have 28 different interests, an empirical error for his personal profile could be approximated by dividing the interests, for example, into seven groups of four interests each and, by finding the mean index of the seven groups and its mean error for each profession. In grouping the interests of an individual, however, it would hardly be fair to choose the sub-groups at random since it is manifestly incorrect to assume that any one interest

[1] The method has been tested later by using a hundred cases, 25 from each group making one of these professions its first choice, and taken alphabetically from students not used in preparing the original table. By this fairer check, 54 per cent. (2.2 times chance) had their highest totals in line with their choices. Seventy-nine per cent showed a total plus record, although not the highest total, in line with their chosen professions.

is of equal value with another for predicting a particular profession. In principle the difference between the largest scores in favor of two vocations might be determined, then the difference between the next largest scores and so on through the 28. A mean difference between these *two* vocations and its error would then be available. By grouping interest scores of comparable weight on the same principle, a mean error of the index for *each* profession can be determined. A fair procedure would seem to be to group the two scores which most definitely indicate a particular profession with the two which predict it least. Then group the next two at each extreme of prediction for that profession and so on. Following this procedure for the particular student cited above, the mean scores of the seven sub-groups for the most favored professions for this student proved to be 12 for engineering and 3 for law, with mean errors of 3.8 and 1.5 respectively. The chances would then be 24 to 1 against his resemblance to this engineering group over that to the law group being reversed. By taking proper representative groups as standards we could thus be reasonably certain whether the differences between the indices for a particular individual were large enough to be significant so far as the comparison of his profile with these standards is concerned.

The evaluation of the total profile of an individual to determine his resemblance to different groups is a problem which promises to require more attention with the increasing study of special traits. In the preliminary stage of such studies a more convenient method than partial correlation is often desirable. The method above suggested is simple and convenient, at least for rough work in classifying individuals. It consists of two steps: First, the evaluation in terms of *sigma* of the percentages of a random group showing different degrees of any trait on the assumption of normal distribution; second, the evaluation of a typical group of individuals on the basis of the difference between that type and the random group for different degrees of the trait in question. The procedure is illustrated in this paper in connection with a vocational guidance problem of evaluating the work interests of a group of college students.

A NEW INVESTIGATION OF JUDGMENTS ON THE SEX OF HANDWRITING

J. S. KINDER

Pennsylvania College for Women, Pittsburgh, Pa.

In 1910, Dr. J. E. Downey repeated Binet's experiment of testing the validity of determining sex from handwriting by the use of untrained judges.[1] Nothing seems to have been done since then to check or disprove their findings. It was accordingly thought to be a wise thing to repeat the experiment at this time—15 years having elapsed since the Downey investigation. Therefore, the writer, assisted by a class in experimental psychology, repeated the Downey-Binet tests, but not *in toto*.

Both Binet and Downey secured their samples by collecting envelopes addressed to themselves. In the case under discussion samples were collected from 50 young men and 50 young women. The latter were all students in a women's college, while the former were studying in a co-educational institution. This eliminated a great source of variation in the writing due to uneven ages, and wide differences in occupation and practice on the part of the writers. Each of the 100 individuals submitting a sample of his writing was asked to write the same sentence so that the same letters would occur in each specimen, thereby offering a better basis of comparison, consequently each wrote "The dog jumps quickly over the fence after the lazy brown fox." This sentence was chosen because it contains every letter of the alphabet.

After the 100 samples had been collected, they were given code numbers and arranged in irregular order. Judgments of these samples —as to whether the writer was male or female—was obtained from 20 judges (sophomores and juniors in a women's college). This gave 2000 judgments, and as none of the judges claimed to have recognized any of the writing or omitted passing opinion on any sample, the 2000 opinions were accepted as *bona fide*.

One thousand three hundred seventy-eight of the 2000 judgments, or 68.4 out of every 100, were correct as to the sex of the writer.

Percentages of correct responses for each judge are given in Table I, as also those given by Binet and Downey.

---

[1] Downey, J. E.: Judgments on the Sex of Handwriting. *Psychological Review*, Vol. 17, pp. 205–216.

TABLE I.—CORRECT JUDGMENTS (EXPRESSED IN PER CENTS) AS TO SEX OF AUTHOR OF A SAMPLE OF HANDWRITING AS FOUND ON THREE EXPERIMENTS

Judges	Correct judgments as found in this investigation	Correct judgments as found by Downey[1]	Correct judgments as found by Binet[2]
1	58	60	65.9
2	59	60	66.4
3	64	61.5	67
4	64	64	68
5	64	66	69
6	65	66.5	69.3
7	68	68	72.9
8	69	68.5	73
9	69	70	73
10	70	70.5	73
11	71	71.4	
12	71	71.5	
13	71	77.5	
14	71		
15	71		
16	71		
17	72		
18	72		
19	72		
20	76		
Average.....	68.4	67.3	63.7

[1] *Ibid.*, p. 207.
[2] Quoted by Downey.

Conclusions are self-evident. The results of the three investigations are quite similar. There is a minor difference noticeable in the variability of the judges. The range in Binet's test is from 65.9 to 73 or 7.1 per cent, Downey's from 60 to 77.5, or 17.5 per cent, and this investigation from 58 to 76, or 18 per cent. Averages of correct judgments are 63.7, 67.3, and 68.4 per cent, respectively (Table I). Fifteen years have not made untrained judges any wiser in determining the sex of a writer from his or her handwriting.

Furthermore, there seems to be no foundation for the argument that present methods of teaching penmanship tend to obliterate individuality and reduce the writing of everybody to a dead level. Most of the men and women who submitted samples for this investigation have

at one time or another come under the influence of one or more of the standardized penmanship methods such as the Palmer, Miller, etc. In spite of this fact the judgment of untrained judges is as good as it was in past years when standardized methods of teaching writing

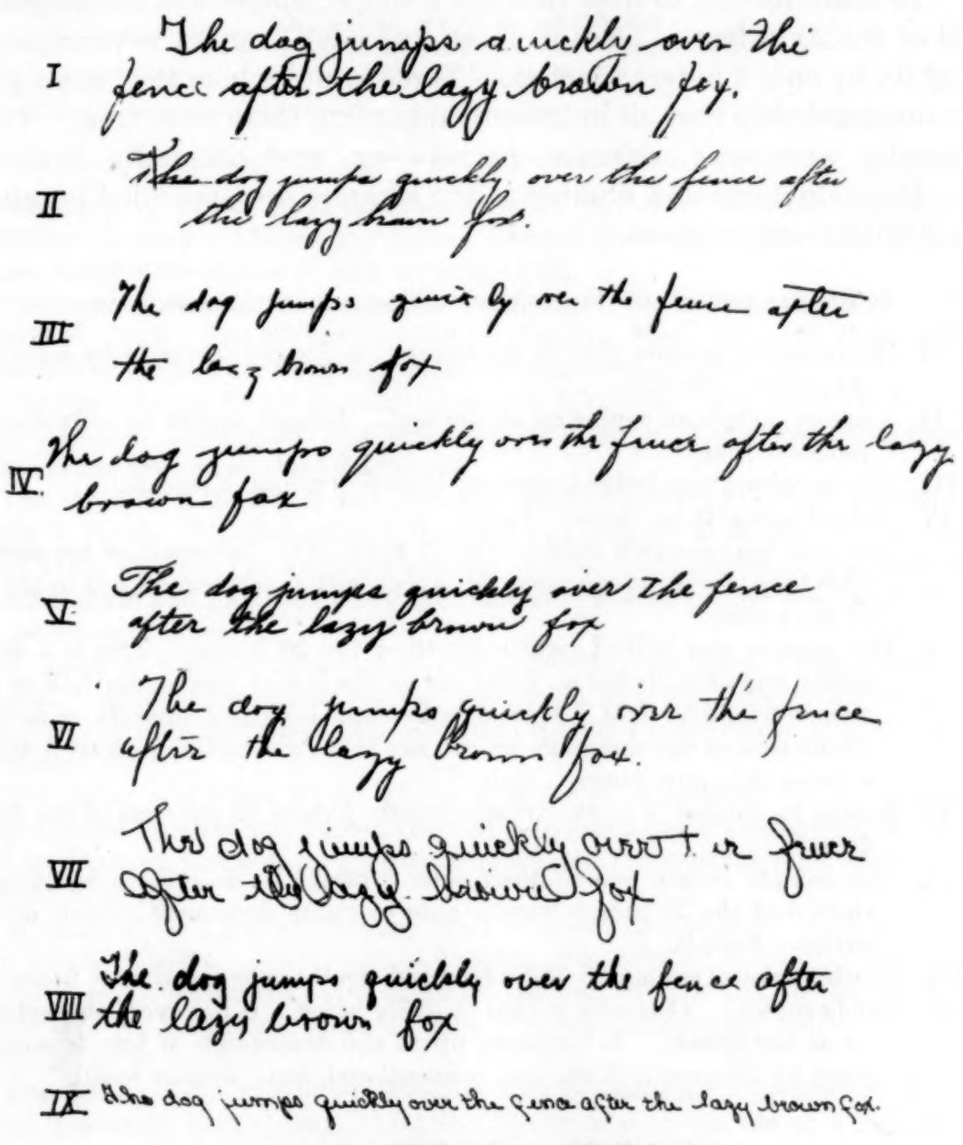


FIG. 1.—(Reduced to one-half size.)

were less used. One is, of course, unable to say that 15 years from now there will be no noticeable effects of present organized methods of teaching writing towards uniformity of skill and design by the sexes.

Miss Downey found that female writing was most frequently erroneously judged, but this investigation gave opposite findings. Perhaps,

the explanation is forthcoming when we remember that the judges were all students in a women's college. This being the case, it is possible that if the "signs of sex" were obscured in a given sample they had a tendency to allocate it to females.

It is interesting to note that not a single sample was misjudged by all of the 20 judges. The two most confusing samples were adjudged rightly by only 3 judges for each. Three samples bore the "sex-signs" so unmistakably that all judgments regarding them were true. These samples were neat, delicate, painstaking, and decidedly feminine.

Reproductions of a number of the samples are appended for study and critical observation.

EXPLANATIONS OF PHOTOGRAPHIC SAMPLES OF HANDWRITING

I. Representing sample that is confusing. Adjudged correctly by only 3 of 20 judges. Female.

II. Another sample as confusing as the first. Judged rightly by only 3 of 20 judges. Male.

III. This specimen was judged correctly only four times. Female.

IV. Judged correctly six times. Male.

The first four samples illustrate what Binet calls "inversion of sex-signs." This type of writing can never be judged with much accuracy as to the sex of the writer.

V. This sample was judged rightly by 10 of the 20 judges. This is a 50-50 rating and as such has no value for by the law of probability half of the guesses would be right. Whenever a sample is judged correctly more than 50 per cent of the time then we can say there is something about it which is more than pure guess. Male.

VI. Similar to number V in that it was rightly judged 50 per cent of the time. Female.

VII. This sample is one which Binet calls typical. It is seldom misjudged. Only 3 of the 20 judges were unable to righly determine the sex of the writer. Female.

VIII. Another typical example. Adjudged incorrectly only one time. Male.

IX. Truely typical. Correctly judged by every judge. It has every characteristic of the female. It measures up to the description of female writing given by Downey: "Colorless, conventional, neat, usually small."

NOTES ON ARTICLES IN EDUCATIONAL PSYCHOLOGY IN CURRENT ISSUES OF OTHER MAGAZINES

REPORTED BY C. O. MATHEWS
Graduate Student, Teachers College
Columbia University

Psychology of Learning and of School Subjects

Summary of Reading Investigations. (July 1, 1924, to June 30, 1925.) I. William S. Gray. Elementary School Journal, Feb., 1926, 449–459. This is an annotated bibliography of such investigations.

Summary of Reading Investigations. (July 1, 1924, to June 30, 1925.) II. William S. Gray. The Elementary School Journal, Mar., 1926, 507–518. Brief summaries of the findings of some of these investigations are presented.

The Psychology of Forced Reading. Grace C. Williams. The Journal of Educational Method, Mar., 1926, 289–294. This is a study of the reading interests shown in the original stories of children.

The Relation of Extensive and Intensive Reading to Permanency of Retention. Carter V. Good. The Pedagogical Seminary and Journal of Genetic Psychology, Mar., 1926, 43–49. The extensive reading groups were superior in the retention of information and problems while the intensive reading groups were superior in the retention of ideas.

A Study of Sentence Structure in Eighth-Grade Composition. Zilla Elizabeth Wiswall. Elementary School Journal, Feb., 1926, 441–448. Herein are reported the effects of different topics and different types of written work upon sentence structure.

The Effect of Certain Types of Speed Drills in Arithmetic. A. I. Messick. Mathematics Teacher, Feb., 1926, 104–109. In four pairs of classes more speed as well as accuracy was obtained when accuracy was emphasized than when speed was emphasized.

A Study of the Effect of Checking upon Accuracy in Addition. John R. Clark and Leona Vincent. Mathematics Teacher, Feb., 1926, 657–71. "The non-checking group performed more additions per unit of time but got fewer correct answers." An index score giving weights to speed and accuracy shows the checking group to be superior.

Overcoming Some Addition Difficulties. H. S. Ballenger. Journal of Educational Research, Feb., 1926, 111–117. The results of the use of a method of remedial teaching to overcome difficulties in column addition were presented.

The Ability to Spell. Jacob S. Orleans. School and Society, Mar. 27, 1926, 407. The relationship is shown between ability to spell in written composition and in dictated lists of words.

Teacher Versus Student Demonstrations in High School Biology. J. S. Cooprider. School Science and Mathematics, Feb., 1926, 147–155. According to test scores and pupils opinions, teacher demonstrations give better results but not much better than those of the better students when closely supervised.

The Role of Habit in Reasoning. Linus W. Kline and Phoebe King Anderson. School Science and Mathematics, Feb., 1926, 156–167. This is a study of the influence of habit in reasoning with dates.

Limits of Learning by Trial and Error. Joseph Peterson. Journal of Experimental Psychology. Feb., 1926, 45–55. An experimental study in the methods of learning and the distribution of practice.

Diagnosis and Remedial Treatment for Poor Spellers. Paul A. Witty. Journal of Educational Research, Jan., 1926 39–440. The steps taken to determine the nature of the defects and their treatment are set forth.

Improvising Comprehension Ability in Silent Reading. Grover H. Alderman. Journal of Educational Research, Jan., 1926, 11–21. Improvement equivalent to two semesters was made by systematic drill.

The Construction of a Reading Vocabulary for the Primary Grades. Arthur I. Gates. Teachers College Record, Mar., 1926, 625–642. Criteria and methods used for the selection of a vocabulary of 1500 words suitable for the first three grades are discussed.

Mental Measurements

Is Classification by Mental Ages and Intelligence Quotients Worth While? T. L. Torgerson. Journal of Educational Research, Mar., 1926, 171–180. The conclusion is reached that proper classification and sectioning reduces failures and "tends to raise the accomplishment quotient of all pupils to normal maximal efficiency."

Age Distribution in the Grades as a Measure of Mentality. Howard J. Banker. Journal of Educational Research, Mar., 1926, 155–170. This study gives a comparison of the mental- and chronological-age medians per grade in the elementary school.

The Problem of Intelligence. H. Pieron. The Pedagogical Seminary and Journal of Genetic Psychology, Mar., 1926, 50–60. A discussion of the nature and measurement of intelligence.

Intelligence Testing in the Kindergarten. Helen L. Duncklee. The Journal of Educational Method, Apr., 1926, 366–369. A description of the Rhode Island test is given with results on a small number of children.

A Comparison of the Scores of College Freshmen and Seniors on Psychological Tests. Robert S. Ellis. School and Society, Mar. 6, 1926, 310–312. The study relates to the problems of elimination from college, transfer of training and assignment of class marks.

How We Judge Intelligence. An Investigation into the Value of an Interview as a Means of Estimating General Intelligence. E. H. Magson. The British Journal of Psychology, Monograph Supplements, IX, 1926. This is a carefully controlled study of the value of the interview and also of judgments after prolonged acquaintance.

The Relation of Intelligence to Behavior. Ira S. Wile, Mental Hygiene, Jan., 1926, 62–74. A discussion of the factors involved in behavior and their relation to intelligence.

The Intelligence of Italian and Jewish Children in the Habit Clinics of the Massachusett Division of Mental Hygiene. Virginia Taylor Graham. The Journal of Abnormal and Social Psychology, Jan., 1926, 371–376. These races are compared on the Stanford-Binet and a number of performance tests. Intercorrelations are reported for all tests.

On the Relation between Negro-white Mixture and Standing in Intelligence Tests. Melville J. Herskovits. The Pedagogical Seminary and Journal of Genetic Psychology, Mar., 1926, 30–42. This study concludes that in a mixture, it is very doubtful whether those having more white blood are innately superior in intelligence to those having more negro blood.

Relative Intelligence of White and Colored Children. L. D. Lacy. The Elementary School Journal, Mar., 1926, 542–546. The average I.Q's of colored children are considerably lower than those of white children at each age and grade level.

The Mental Capacity of Amercian-Born Japanese Children. Marvin L. Darsie. Comparative Psychology Monographs, Vol. III No. 15, Jan., 1926. A study of the differences in intellectual capacity between the American and Japanese races.

Intelligence and Immigration. Clifford Kirkpatrick. Mental Measurement Monographs, No. 2, Jan., 1926. This is a comprehensive study of intellectual differences in immigrant groups. An excellent review of previous investigations is included.

A Psychological Study of Immigrant Children at Ellis Island. Bertha M. Boody. Mental Measurement Monographs, No. 3, Feb., 1926. Included in this monograph is an experimental study of the abilities of foreign children by means of a number of tests.

The Intelligence of Full Blood Indians. Thomas R. Garth. Assisted by T. J. Serafini and Dewey Dutton. The Journal of Applied Psychology, Dec., 1925, 382–389. The results are based on the measurement of 1102 Indians with the National Intelligence Test. Approximate average IQ's and mental age norms are given for grades four to eight.

A New Point Performance Scale. Grace Arthur. The Journal of Applied Psychology, Dec., 1925, 390–416. The method of obtaining point values for each test and tables for determining point values for any score are given.

Character and Personality

Testing the Knowledge of Right and Wrong. Hugh Hartshorne and Mark A. May. Religious Education, Feb., 1926, 63–76. This article describes the construction and validation of some of the tests being developed by the Character Education Inquiry.

A Critical Study of Rating Traits. A. P. Nanninga. Educational Administration and Supervision, Feb., 1926, 114–119. The data are based upon the ratings of five teachers given to thirty pupils on fifteen traits.

The Permanence of Vocational Interests after Three Years. E. Earl Franklin. School and Society, Apr., 3, 1926, 438–440. Three out of five junior high school pupils maintain the same choice of occupation after three years.

Interests Studied Quantitatively. Ruth M. Hubbard. The Journal of Personnel Research, Jan.—Feb., 1926, 365–378. A report of the measurement of differences between socially and mechanically inclined college students in relation to vocational selection.

The Prediction of a Disparity between Scholarship and Intelligence. J. B. Miner. The Journal of Applied Psychology, Dec., 1925, 357–363. Prediction was attempted by use of Army Alpha scores and the Downey test was applied to the extreme groups.

NEW PUBLICATIONS IN EDUCATIONAL PSYCHOLOGY AND RELATED FIELDS OF EDUCATION

CONDUCTED BY JOHN HOCKETT
The Lincoln School of Teachers College

THE NEED FOR LESS SUBJECTIVE BOOK REVIEWS, WITH AN ILLUSTRATION

JOHN P. HERRING
Teachers College, Columbia University

I offer a second review of Averill: *Elements of Educational Psychology* partly in order to add to the first[1] a certain series of roughly scaled samples from the author's text which were omitted before, and partly in order to reiterate and to amplify an argument at which the earlier review hinted and which a review of Thurstone: *Fundamentals in Statistical Method*[2] openly suggested,—an argument for the application of science to the estimation of values in books.

Unverified opinion is still the vogue. One weighing the relative merits of science and opinion may perhaps always rightfully reserve space for a review expressing the judgment of a single judge who has repute in the field. There is a most valid intrinsic interst in knowing what J. McKeen Cattell thinks of a book by John B. Watson. But there is also reason to create the conventional scientific disclosures and improvements in the validity and objectivity of reviews. Three reviews like Murchison's,[3] Willing's,[4] and mine, differing as they do in aspects selected for comment, in method of treatment and in final estimate, illustrate still further the fundamental disagreements exhibited among rankings by the nine judges of the samples to follow. Half of these judges have national repute or better as teachers of educational psychology and all of them are keen students in educational psychology. The Pearson correlation between pairs of judges ranged from zero to 0.8 or 0.9 and that between four judges and five

[1] Herring, John P.: Averill, L. A.: "Elements of Educational Psychology." Houghton Mifflin Co., Boston, 1924. *Journal of Educational Research*, May, 1925.

[2] A review by John P. Herring, this journal, this number.

[3] Pedagogical Seminary, March, 1925.

[4] This Journal, November, 1925.

judges was about 0.8. If confirmed by better investigations, it would be suggested that competent persons differ in their opinions about books in their own general field as much as they do about men in their own general field—a superlative misfortune, if true, for the space which scientific journals have already devoted to book reviews.

Instead of repeating the details descriptive of the very crude scaling of the samples given below, I refer to my earlier review. The samples range from one so good that all of the nine judges would include it in a text on Educational Psychology, to one so faulty that only one judge would include it. While I have only my own judgment for the contention that the scale justly represents the book, that judgment has perhaps indirect support in a correlation of 0.9 (the highest) between my independent ranking of samples and the average ranking of the nine judges—a situation suggesting a weakness in the scale: perhaps less time spent by judges other than myself. My argument does not now mainly concern the value of the book, which may be taken as worth some sort of average of the 11 opinions including Murchison's and Willing's. Perhaps these men will pardon a suggestion that while their reviews are of course in every way competent, the judgment of the nine has on the one hand better chance of being the more objective and on the other no chance at all of being as objective as science should one day make such estimates. Scientific valuation of manuscripts and books would save money, work, courage and humiliation.

The Scale of Samples

+1.5 SD

All nine judges would include

"Previous to the time of Herbart, psychologists were committed to the notion that the mind is made up of a number of abstract mental powers which were called 'faculties.' Thus, there was the faculty of judgment, of reasoning, of memory, of imagination, if will,[1] of emotion, etc.

The newer psychology has materially changed this . . ."

+1.0 SD

Seven judges would include

"Not memory, but memories. One is often inclined to say of himself, 'I have a poor memory,' or, 'I have a very good memory,'

[1] But see sample—1.7 SD on p. 352.

implying that memory is a general faculty of the mind. The experimental evidence, however, rather upsets this theory. An individual appears to have a different memory for almost everything he retains, and one ought in consequence to speak, not of 'memory,' but rather of 'memories.' In our everyday experience we find evidence of this desparity of memorial capacities in the circumstance that the same individual may have a poor memory for names, but an excellent memory for faces . . ."

+.7 SD

Eight judges would include

"But where would the pupil who got the highest score in reading rank in arithmetic, or in writing, or in any or all the other subjects in the curriculum? Where would the pupil who was poorest in reading rank in these other subjects? Would the middle pupil in reading score mid-way also in them? These questions suggest the meaning of the term *correlation*, by which is meant the relation that exists between two functions. If the best pupils in English are found to stand among the best pupils in arithmetic, for example, we say there is a high positive correlation between ability in these two subjects; if the poorest pupils rank similarly with the poorest, and the middle with the middle, we say again, there is a high positive correlation between ability in these subjects. In other words, when existence of one trait is accompanied invariably by existence of another trait in reasonably equivalent amounts, the correlation is a positive one. If, on the other hand, it should develop that those pupils who rank highest in reading rank lowest in arithmetic, the correlation is said to be negative. In other words, when two traits tend to be mutally exclusive, the correlation is a negative one.

Loose phrasing.

+.4 SD

Four judges would include

"In this lesson, continuing our attention to the matter of skilful motivation of school work, we purpose to make further analysis of it from the viewpoint of its reactions upon the minds of the pupils. Some of the questions we shall be concerned with attempting to answer are: Do the feelings play any appreciable part in the learning process? What is the value of pleasurable accompaniments of the learning process? Of positively disagreeable or annoying accompaniments? What is the effect upon the pupils' minds of a 'colorless' lesson? When

a child is studying a lesson, for example in geography, is he doing anything else that is bound to make him a different personality thereafter.

+.3 SD

Five judges would include

Differences between animal and human learning:

"Animal learning is limited to a very restricted range of sensori-motor responses."

"The absence of language is an important factor in the delimitation of the animal's possible range of acquisition."

"The animal learner is unable to abstract from concrete experiences general principles which will help in subsequent experiences."

"Imitation, which plays an important rôle in human learning, is of relative insignificance as a method in animal learning."

"The plateau phenomenon, often found in human, does not exist in animal learning."

+.3 SD

Six judges would include

"The inevitableness of this tendency to connection will depend ordinarily upon the number of time the neuromuscular mechanism has been linked up in that definite order."

(Other factors?)

+.2 SD

Four judges would include

"Nor does the essence of the meaning of education lie in that narrow, utilitarian interpretation of it that views it largely in a bread-and-butter sense; rather it lies in the infinitely broader but still utilitarian interpretation which sees in education a means of stimulating and equipping the individual to prepare for the fullest possible living—industrial, intellectual, family, social, spiritual."

(Psychology or philosophy? Education as preparation for future living merely?)

+.1 SD

Six judges would include

"Save in a subject as exact as arithmetic, there is certain to be a wide range of difference between the standards of any two or more persons who are grading papers or calculating the quality of classwork."

(Exception valid?)

−.7 SD

Three judges would include

"It is also true that an idea, as well as a perception, may initiate habitual motor reaction. Thus, sight of the toothbrush hanging over the wash stand (perception) may impel me to brush my teeth; or again, the thought that it is my usual time for brushing them (idea) may equally impel me to do it. Where such motor response is initiated by an idea rather than by a perception, we ordinarily refer to the action as *ideo-motor* action."

(Antequated psychology.)

−.9 SD

Three judges would include

"It represents racial man in the present child. As we stated above, valuable experiences which the unknown ages have brought into the life of the race are carefully preserved by Nature in the nervous system....."

"..... the innate tendencies of Nature's child—the inevitable ontogenetic product of cumulative ages of phyletic accretions—are found by every teacher and every parent to be strong and insatiable. In the child's nervous system are emblazoned all the diverse records of the century-old struggle for survival. And that is why children crave activity, adventure, possession, companionship, and all the rest."

(Antequated psychology.)

−1.4 SD

Two judges would include

"The plateau phenomenon, often found in human, does not exist in animal learning.

−1.7 SD

One judge would include

"Subject W, on the other hand, is by no means actuated by a determined will. Her will functions only intermittently."

(Faculty psychology.)

−1.9 SD

Two judges would include

"There are a considerable number of tests of learning capacity on the market. Probably a combination of at least two will prove the

most satisfactory: a substitution test, and a logical memory test. Pyle recommends the employment of two additional ones: card-sorting and nonsense syllables. It is probable, however, that a substitution test and a logical memory test, properly handled, will suffice for the purpose. It is obvious, of course, since learning is associating new data with one another, that any test devised by the teachers may be used as a valid measure of learning capacity, provided the data to be learned have not previously been experienced."

(This paragraph contains a truly remarkable statement which seems to say that it follows (sic) from the very definition of learning, that "the" teachers are competent to construct "valid" tests of learning capacity!)

JOHN P. HERRING.

Teachers College, Columbia University.

A TEXT IN EDUCATIONAL STATISTICS

The Fundamentals of Statistics, by L. L. Thurstone. New York: The Macmillian Co., 1925. Pp. XVI + 237.

Such a title as: *Fundamentals of Statistics in Education* might somewhat more specifically suggest the content of Thurstone's *The Fundamentals of Statistics*, for although the illustrative material includes some physics and many problems with abstract numbers, the most coherent pertinence seems to be the successive bearing of each of a number of special mathematical fields upon the measurement of intelligence and learning: variables and their constants; coördinate geometry; binomial expansion and probability; and correlation. The statistics presented, while indeed concretely fundamental to educational measurement, would, for instance, hardly serve as *the* fundamentals of statistics for economic measurement, in which correlation, regression, and standard error have the special treatment, interpretation and illustration required in determining the accuracy of index numbers, in eliminating seasonal from saecular fluctuation, in measuring lag, in studying the intercorrelations among time series, and in predicting futures. Thurstone's synthesis of contents, illustrations, graphs and problems, with certain interests of students, combining clarity with precision in its expression, seems highly useful as a text when a teacher wishes to teach just this content, and as a supplementary book, when an elementary student wishes to fill certain interstices

in the statistical structure he is building from other sources. In view of the natural interest of college students in the meaning and criticism of intelligence examinations when used to measure themselves and their kind, Thurstone has proceeded strategically to illustrate first and often, by means of data, distributions, means, sigmas, and correlations representing the intelligence of groups in college. The impression that it might have been also strategic, consistently with the topics actually treated, to replace many, perhaps all, of the abstract problems with those which are concrete, *e.g.*, in college education; or, consistently with the actual title of the book and at the same time perhaps with some of the interests of some of the students, by problems in part from such as biology (intelligence and cranial capacity), medicine (prediction from measured symptoms), economics (prediction of prices), sociology (analysis and prediction of population), anthropology (criteria for distinctions between races), should be offered with the recollection that both the content and the problems are matters that have survived seven years of students' responses. My parenthetical arm-chair suggestions of concrete problems may be mostly the wrong ones; only some sort of objective method can really tell either what is best to include in a book or what a reviewer should say.

Comparison of any text with a random sampling of others nearly related suggests the need for research on the selection of content. The binomial expansion for example, may be intrinsically interesting to know, to teach and to learn; it may occur in one but not in the other of two texts designed for the same ends and the same students. What, however, would be the regression weight representing the relative contribution of any topic to the objectives of a course, and in turn of objectives to more inclusive purposes? While my opinion in this instance is with the author, I should on the one hand, like to know much more objectively, and, on the other, I should be astonished to find the content of more than a minority of the thousands of contemporaneous texts largely justified. Despite the values and economies which could easily be imagined as outcomes of organization for the objective valuation of publications, the venerable institution of the book-review seems still a stronghold of widely divergent opinion.

If even college teachers may rightly allow a changing philosophy of education to mould a changing methodology, books like this might conceivably be written with the purposes of survey and solutions—survey of students' problems selected because of vital import and susceptibility of statistical treatment; and solutions of the problems

thus disclosed, by means of the statistical methods appropriate. This concept would turn many a book another side out. While no such idea has yet dominated any college text in educational and mental measurement, it is my hypothesis that the plan would be worth scientific valuation along side of the older one in which, although interest may be thought to be present as a means, the real outcome is too often merely the learning of techniques.

The book adds to our stock many a detail of exposition needed by the beginner for an easier understanding of mathematical concepts which have been difficult because usually expressed in the eliptically foreign tongue of mathematical symbols. The movement toward easier exposition in this generation has both motive and justification in the unparalleled growth of college attendance, with its almost inevitable easing of standards of admission. If college students are ever generally selected so well that their quality is maintained despite their growth in numbers, they may inherit from this period of transition a collection of educational aids as clear as this book of Thurstone's—a combination, of intellect and means, which would accelerate learning.

JOHN P. HERRING.

Teachers College, Columbia University.

TABLES OF PROBABLE ERROR

Tables of the Probable Error of the Coefficient of Correlation as Found by the Product Moment Method, by Karl Holzinger. Cambridge University Press, 1925. Pp. 36. Price 3s 9d net.

This is one of the series of *Tracts for Computers* edited by Karl Pearson, and contains an introductory note by the editor explaining the use of the tables. Two tables are given. Table A gives the probable error to four decimal places for all values of N from 20 to 100 by units, and for values of r from .00 to 1.00 by increments of 0.2. Table B gives PE for values of N from 100 to 1000 by tens and for values of r as in Table A. Tables of proportional parts are given to facilitate interpolation.

These tables will be of service to all persons using the product moment coefficient of correlation, and especially to the large number of students and teachers who do not possess computing machines.

J. H.

A Comprehensive Summary of Abnormal Psychology

Outline of Abnormal Psychology, by William McDougall. New York: Charles Scribner's Sons, 1926. Pp. XIII + 566.

"I regard this volume" says the authors, "as an early, if not a first, attempt to bring into something like systematic order a very large and difficult province of science, hitherto cultivated in patches rather than as a whole." Bringing to the task his distinctive viewpoint in academic psychology, his preparation in medicine, five years of war experience in handling neuroses, his experience in being psychoanalyzed by Dr. C. G. Jung, and his strong interest in psychicalres earch and metaphysics, the author has brought together in a single volume a rich and varied assortment of problems and doctrines.

Among the schools of abnormal psychology the author gives particular attention to Jung, Freud, Adler, Stekel, and Janet, classing himself among those styled "the school of integral psychology," and including beside himself, Dr. Morton Prince, W. H. R. Rivers, J. A. Hadfield, Bernard Hart, Crichton Miller, William Healy, and others. Of Freud he says, "I believe that Prof. Freud has done more for the advancement of psychology than any student since Aristotle" Especially the dynamic, hormic, character of Freud's approach appeals to the author. "The fundamental notion guiding Freud's therapy is correct, namely, the notion that exploration of the mental nature and origins of the disorder and the revelation of them to the patient is the most essential step in therapy." At the same time the author is keenly sensitive to the inadequacies of Freudianism, the pansexuality, the symbolism, the ill-defined Ego-instincts, the disagreements between cases and theories, the mythology, and the many hopelessly broad generalizations on the basis of slender evidence. His summary of Jones' Freudian approach to fear is delightful. "All fear is morbid anxiety; it is the Ego's defensive reaction against the claims of unrecognized sexual hunger; and fear, the commonest neurotic symptom, is fear of 'the Unconscious,' and is constructed by 'the Unconscious' in order to protect the Ego against fear." McDougall stands almost alone among the psychologists who have so far deigned to discuss Freud, in having read enough of Freud to know that there have been many significant changes of recent years, in the Freudian theories. Even in this volume, the author presents rather length discussion of the hedonistic "pleasure principle," the supposedly sexual character of all neuroses, the Oedipus complex in adults, the method of therapy

by talking out the infantile experiences, although as he recognizes at the end of each discussion, none of these theories are now sustained by Freud.

Adler, the author believes, has provided a good supplement to Freud, because of the Adlerian emphasis upon the importance of the Ego-instincts. Jung is regarded as having gone furthest in the development of a desirable emphasis in psychoanalysis. Yet "when we are confronted with such a theory as Jung's Collective Unconscious, a theory which raises the deepest questions of biology and philosophy, it behooves us to go slowly, though with open minds."

Beginning with an emphasis on the functional character of many disorders, and the uselessness of many traditional classifications, the author next presents a study of fatigue, drugs, and sleep. Hypnotism is regarded as undoubtedly the most important, the most fruitful and far-reaching, method of experimental psychology and several theories of hypnosis are analyzed. There follow several chapters on dreams, amply illustrated. Then such disorders as complexes, dissociation, automatisms, fears, anxieties, regression, tics, compulsions, sex perversions, delusions, and hallucinations receive examination from several points of view. Then the psychoses appear under the titles "Exaltation and Depression," "Schizophrenia," and "Epileptoid Seizures." The next section deals with the systems of therapy employed by the various groups of psychoanalysts, concluding with the author's synthesis. "In all psychotherapy there are two essential steps: first the process of exploration by which the nature and origin of the morbid state are as far as possible brought to light and made clear to the patient; secondly the process of readjustment of the patient's mental life, more especially of his affective tendencies."

The concluding chapters present the classical cases of alternating and dissociated personalities, leading to a discussion of the real nature of personality. The author's theory of character is that "the only sentiment which can adequately fulfil the function of dominating and harmonizing all other sentiments is the sentiment of self-regard, taking the form of a self-conscious devotion to an ideal of character." Further, viewing integration from the point of view of consciousness rather than from such a behavioristic point of view, the author suggests that normal human personality is a hierarchy of "monads." "The obvious, and I believe, inevitable inference from the facts, is that I who consciously address you am only one among several selves or Egos which my organism, my person, comprises. I am only the

dominant member of a society, an association, of similar members. There are many purposive activities within my organism of which I am not aware, which are not my activities but are those of my associates. I am conscious at any moment only of those processes within the organism and of those impressions from without, which it is most necessary that I should take cognizance of." The communication between these monads is thought to be telepathic.

The author finds in his wide examination of problems, no evidence which should cause him to alter the theory of instincts and of physical processes, set forth in his "Social Psychology." Most of the criticisms which have been directed at that work, will be found to be equally applicable here.

The reader will find the volume useful not only as a further elaboration of a widely criticized viewpoint in psychology, but as one of the best, perhaps the best, of the existing summaries of various theories regarding those fascinating and unsolved problems which lie far out on the border line of psychological knowledge. The book is an excellent stimulant for any smug security which may have resulted from too long confinement in the limited area of simple and measurable reflex-type bonds. Of particular interest are the 60 cases and many more illustrations, selected to include such classic cases as Rivers' claustrophobia, Brill's case of the dream about Liconor Bay, and Prince's study of the many personalities of Miss Beauchamp, as well as many extraordinary cases which have arisen in the wide experience of the author.

GOODWIN B. WATSON.
Teachers College, Columbia University.

THE TEACHER

Personnel Problems of the Teaching Staff, by Ervin Eugene Lewis. New York: The Century Company, 1925. Pp. XVII + 460.

This book gives a conspicuously well organized presentation of available fact and current discussion on the complicated question of teacher management. Its 22 chapters deal with as many important topics, such as recruiting, teacher-training, certification, selection, home talent, married women teachers, teacher measurement, salaries, turn-over, failure, and organization. Each chapter gives compactly and with much use of tabular and catalogue forms of statement the results of investigations, experiments, practical experience, and expert

suggestion about the particular problem in question. Although the details covered are multitudinous, the organization, style, and manner of summarizing make the book easy to read. It appears to be designed to satisfy equally the needs of the graduate student of educational administration and the harried superintendent looking for condensed accounts of ways to get himself out of personnel predicaments. It is true that many more problems are recognized than there are solutions to match, but the bringing together of so much that is practically suggestive constitutes a very fine educational service.

The point of view is wholly in keeping with that of the most advanced personnel work in industry and business. It identifies the major problem as one of making "each teacher-in-her-work unit as effective as possible." This is not to be accomplished through mechanical means, though these may be of assistance; it will be done partly through the intelligent and sympathetic personal attention of all charged with the supervision and administration of education. It will be done partly by the teachers themselves in associations for the improvement of their own profession standing and personal efficiency. It will never be done completely until the people who support the schools realize the meaning of modern education and appreciate the work of those carrying it on. This volume by Superintendent Lewis is excellently conceived to impress these fundamental truths.

M. H. Willing.

Problems of the Rural School

The Administration of Consolidated and Village Schools, by John C. Almack, and James F. Bursch. Boston: Houghton Mifflin Company, 1925. Pp. XV + 466

Problems of organization and administration of consolidated and village schools are described. The "job analysis" study in many instances reveals unexpected content for analysis. The analysis of the problem of transportation is a fine example.

The authors state the similarity of city and rural school administration in so far as the problems of supervision of classroom instruction are concerned. They proceed to show that the problems of organization and administration in city and rural schools are in the main different. These differences are found in the degree of specialization of the functions that are performed. The rural administrator is a "jack of

all administrative functions" while a city school administrator usually delegates certain functions to specially trained subordinate officers.

However true it may be that many city, county and State school administrators are recruited from the "training ground" of rural school administration, it does not follow that such recruits are professionally trained for the administrative problems of the school systems of city, county or state. The complex and specialized nature of the content of city school administration problems, for example, compared to the simple and general nature of the content of rural school administration problems would hardly admit the rural "training ground" as a reliable or valid place in which to prepare the modern professionalized city school administrator. City school administration demands an expert in handling the problems of the public school. Evidence indicates that many city school administrators who have been trained through rural experience realize the value of professional ized training for their work as an administrator of a city school system. The rural or even the city "training ground" is not necessarily the best or only place to prepare professional city school administrators.

Better administration in village and consolidated school systems can probably be attained through professionalized courses of study assisted by such works as the above volume. Likewise better administration in city, county and State public school systems can probably be realized through professionalized courses of study assisted by carefully assembled works containing differential analyses of their more specialized problems and issues.

The above volume is a beginning in a field in which much research is needed. It seems that a vew volumes should be assembled which would display an inventory of the jobs and problems in each phase of rural school administration much in the same way as has recently been done in city school administration. CHARLES C. WEIDEMANN.

SEP 21 192

cop 2